Brain Network Dynamics in Deviance Response and Auditory Perception

DISSERTATION

zur Erlangung des akademischen Grades Doktoringenieur
Dr.-Ing.

vorgelegt der Fakultät für Informatik und Automatisierung
der Technischen Universität Ilmenau

von M.Sc. Shih-Cheng Chien

1. Gutachter: Prof. Dr. habil. Thomas R. Knösche
2. Gutachter: Prof. Dr.-Ing. habil. Jens Haueisen
3. Gutachter: Prof. Dr. Marc Schönwiesner

Abstract

Neural responses to sudden changes can be observed in many parts of the sensory pathways at different organizational levels. For example, deviants that violate regularity at various levels of abstraction can be observed as simple On/Off responses of individual neurons or as cumulative responses of neural populations. The cortical deviance-related responses supporting different functionalities (e.g., gap detection, chunking, etc.) seem unlikely to arise from different function-specific neural circuits, given the relatively uniform and self-similar wiring patterns across cortical areas and spatial scales. Additionally, reciprocal wiring patterns (with heterogeneous combinations of excitatory and inhibitory connections) in the cortex naturally speak in favor of a generic deviance detection principle.

The proposed generic deviance detection principle separates the generation of deviance response into two functional stages: regularity formation and change detection. The principle suggests that the deviance-related responses observed in the cortex, such as the cortical On/Off responses, the cortical omitted-stimulus response (OSR), and the mismatch negativity (MMN), can be regarded as change responses at different levels of abstraction.

The network model based on the principle reproduce several experimentally observed properties of cortical deviance-related responses, which include the diverse temporal profiles of On/Off responses, the linear relationship between OSR latency and input stimulus onset asynchrony (SOA), and the slow and fast responses in the sequence MMN paradigm. Regarding change detection, the simulation results suggest that the emergence of change detectors relies on the involvement of disinhibition. An analysis of network connection settings further suggests a supportive effect of synaptic adaptation and a destructive effect of N-methyl-D-aspartate receptor (NMDA-r) antagonists on change detection. Regarding regularity formation, the simulation results suggest the need for a place coding scheme, a larger

time constant of inhibitory population, and short-term plasticity to support a steady neural representation of regularity.

Several model predictions are provided for experimental validation. First, the cortical deviance-related responses might show similar laminar profiles, especially the activity of the inhibitory neurons where the change detection takes place. Second, the NMDA-r antagonists would generally dampen the cortical Off response, the cortical OSR, and the MMN. Third, there might be distinct geometric distributions of change detection and regularity formation, since the two functions emerge from different network properties such as the time constants and connection patterns.

This thesis provides a new viewpoint on the neural mechanisms underlying the generation of deviance response. Future research topics, such as the attention mechanism in perception, the functional roles of various types of inhibitory neurons, and the process of higher cognitive functions such as language acquisition and comprehension, can be based on the current implementation of simulations.

Kurzfassung

Neuronale Reaktionen auf plötzliche Veränderungen des sensorischen Inputs können in vielen Teilen der sensorischen Bahnen auf verschiedenen Organisationsebenen beobachtet werden. So können beispielsweise Abweichungen, die die Regelmäßigkeit auf verschiedenen Abstraktionsebenen verletzen, als einfache Ein-/Aus-Reaktionen einzelner Neuronen oder als kumulative Reaktionen neuronaler Populationen beobachtet werden. Aufgrund des relativ einheitlichen und selbstähnlichen Verdrahtungsmuster im Kortex scheint es unwahrscheinlich, dass die verschiedenen kortikalen Reaktionen, die unterschiedliche Funktionalitäten unterstützen (z.B. Lückenerkennung, Chunking, etc.), jeweils auf spezialisierten kortikalen Verschaltungsmustern beruhen. Darüber hinaus sprechen reziproke Verdrahtungsmuster (mit heterogenen Kombinationen von exzitatorischen und inhibitorischen Verbindungen) im Kortex für ein generisches Prinzip zur Erkennung von Abweichungen.

Das vorgeschlagene generische Prinzip der Abweichungserkennung unterteilt die Erzeugung der Abweichungsreaktion in zwei Funktionsschritte: Regularitätsbildung und Änderungserkennung. Das Prinzip legt nahe, dass die im Kortex beobachteten Reaktionen, wie die kortikalen Ein/Aus-Antworten, die kortikale Auslassungsreaktion (OSR) und die Mismatch-Negativität (MMN), als Änderungsreaktionen auf verschiedenen Abstraktionsebenen betrachtet werden können.

Das Netzwerkmodell, das auf diesem Prinzip basiert, reproduziert mehrere experimentell beobachtete Befunde, zu denen die unterschiedlichen zeitlichen Profile der Ein/Aus-Antworten, die lineare Beziehung zwischen OSR-Latenz und Input Stimulus Onset Asynchrony (SOA) und die langsamen und schnellen Reaktionen im Sequenz-MMN-Paradigma gehören. In Bezug auf die Erkennung von Veränderungen deuten die Simulationsergebnisse darauf hin, dass für das Auftreten von Verän-

derungsdetektoren ein Disinhibitionsmechanismus erforderlich ist. Eine Analyse der Verbindungsstärken im Netzwerk deutet weiterhin auf einen unterstützenden Effekt der synaptischen Anpassung und einen destruktiven Effekt von N-Methyl-D-Aspartat-Rezeptor- (NMDA-r)-Antagonisten auf die Änderungserkennung hin. In Bezug auf die Regularitätsbildung deuten die Simulationsergebnisse auf den Notwendigkeit für ein raumcodierenden Schema, eine größere Zeitkonstante der hemmenden Population und kurzfristige Plastizität hin, um eine stetige neuronale Repräsentation der Regularität zu unterstützen.

Für die experimentelle Validierung können wir mehrere Vorhersagen aus dem Modell ableiten. Erstens sollten die verschiedenen kortikalen Abweichungsreaktionen ähnliche laminare Profile aufweisen, insbesondere bzgl. der Aktivität der inhibitorischen Neuronen, in denen die Änderungserkennung stattfindet. Zweitens würden die NMDA-r-Antagonisten im Allgemeinen die kortikale Aus-Reaktion, die kortikale OSR und die MMN dämpfen. Drittens könnte es unterschiedliche räumliche Verteilungen der Änderungserkennung und Regularitätsbildung geben, da die beiden Funktionen aus unterschiedlichen Netzwerkeigenschaften wie Zeitkonstanten und Verbindungsmustern entstehen.

Diese Arbeit bietet einen neuen Blickwinkel auf die neuronalen Mechanismen, die der Detektion von Abweichungen zugrunde liegen. Zukünftige Forschungsthemen, wie der Aufmerksamkeitsmechanismus in der Wahrnehmung, die funktionelle Rolle verschiedener Arten von hemmenden Neuronen sowie höhere kognitive Funktionen wie Spracherwerb und -verständnis, können auf der aktuellen Implementierung des Modells basieren.

Schlüsselwörter: auditive Wahrnehmung, kortikale Berechnung, neuronale Massenmodellierung, Disinhibition, kurzfristige Plastizität, E/I-Gleichgewicht, NMDA, Ein/Aus-Reaktionen, kortikale Auslassungsreaktion, mismatch negativity

Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen:

1. Prof. Dr.-Ing. habil. Thomas R. Knösche: Interpretation of the analyses, Diskussion des vorgeschlagenen generic deviance detection principle

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt. Ich bin darauf hingewiesen worden, dass die Unrichtigkeit der vorstehenden Erklärung als Täuschungsversuch angesehen wird und den erfolglosen Abbruch des Promotionsverfahrens zu Folge hat.

Leipzig, 25. September 2019   Shih-Cheng Chien

Acknowledgement

I would like to express my deepest gratitude to my supervisor Thomas Knösche for his guidance and patience. His flexible yet rigorous thinking has been inspiring, and the ideas he brought into our discussions have been invaluable. I would also like to thank my advisors Burkhard Maess and Marc Schönwiesner for providing helpful suggestions in our meetings, which allowed me to find connections between theories and experimental observations. Without my supervisor and advisors, I would not have been able to explore and dwell on my favorite research topic freely.

It has been a delightful experience working at the Max Planck Institute for Human Cognitive and Brain Sciences (MPI CBS) in Leipzig. From there, I have had the pleasure to meet and work with many lovely people. I am particularly grateful for the support provided by Veronika Krieghoff and Susann Glasewald, the coordinators of the institute. Their assistance in the administrative routines has been of great help to me. I would also like to extend my thanks to my former and current colleagues of the institute for the inspiring discussions and various forms of support: Daria Goranskaya, Seung-Goo Kim, Chiao-Yi Wu, Peng Wang, Tim Kunze, Hermann Sonntag, Pei-Cheng Shih, Jae-Hyun Cho, Vincent Cheung, Konstantin Weise, Richard Gast, Ruxue Gong, Helmut Schmidt, Alejandro Tabas, Pei-Ju Chien, Tatsuya Daikoku, Ayaka Tsuchiya, and Katarzyna Gugnowska.

I would like to thank Venerable Master Ri-Chang and Venerable Yuan-Li for their teaching and guidance. The wisdom gained from them has helped me find inner peace in tough times. I would also like to thank my supportive friends, especially Ming-Ying Chuang, Hsiu-Hsiang Chen, Daniel Craven, Joseph Feng, Tsai-Chin Kuo, Shu-Ying Wang, Shu-Fen Lee, Sandy Chang, Judy Chen, Sara Wang, Shao-Yi Huang, Mark Hamilton, Julia Schmidt, and Shau-Wen Lin, for enriching my knowledge and my life.

Abbreviations

Abbreviation	Meaning
A1	Primary auditory cortex
EEG	Electroencephalography
ERP	Event-related potential
EPSP / IPSP	Excitatory / inhibitory PSP
fMRI	Functional magnetic resonance imaging
FRF	Frequency receptive field
ISI	Inter-stimulus interval
IOI / SOA	Inter-onset interval / stimulus onset asynchrony
LAN	Left anterior negativity
LFP	Local field potential
MEG	Magnetoencephalography
MMN	Mismatch negativity
NMDA	N-methyl-D-aspartate
NMDA-r	N-methyl-D-aspartate receptor
OSR	Omitted-stimulus response
PSP	Postsynaptic potential
RMS	Root mean square
ROI	Region of interest
STP	Short-term plasticity
STDP	Spike timing dependent plasticity
TWI	Temporal window of integration

Contents

Chapter 1

Introduction

"What I cannot create, I do not understand."

– Richard P. Feynman

1.1 Motivation

The motivation rests on the pursuit of *a unifying theory of perception.* A theory helps to answer the *how* and *why* questions. By attributing functional roles to the observed phenomena (e.g., neural activity) and investigating the functional interaction, we may have a chance to understand the neural mechanisms that underlie perception.

Perception is about finding features and patterns. This includes finding a deviant feature out of standard ones. I chose to investigate *deviance response* as a first step in studying perception for several reasons. First, the deviance response can be observed by different recording techniques (e.g., EEG/MEG, LFP, Calcium imaging, etc.) across various modalities (e.g., visual, auditory, olfactory, gustatory, and tactile) and species (e.g., primates, rodents, reptiles, fish, etc.). Therefore, an extensive database and clues are available for refining a theory. Second, deviance response serves several functional roles, which can include energy saving (by data compression), features extraction (e.g., chunking in speech, beat perception in music, etc.), and attention switching. These functional roles connect broadly to other research topics in perception. Third, a better understanding of deviance response

benefits the clinical applications (e.g., for diagnoses and treatment of patients with schizophrenia, dyslexia, etc.).

1.2   Aim and contributions

The aim is to investigate the neural mechanisms underlying the generation of deviance response, and to associate the neural mechanisms with other phenomena in perception. To achieve this, the related experimental and theoretical studies were reviewed, and a simulation approach was used to bridge the hypotheses and experimental observations. The main contributions of the thesis are listed below.

1. **A new hypothesis was proposed.** There have been several hypotheses proposed to account for the generation of deviance response, and there are several conflicts and confusions among these hypotheses. A *generic deviance detection principle* was proposed in an attempt to reconcile the conflicts and confusions. The proposed principle accounts for the different *cortical deviance-related responses* with a similar underlying neural mechanism. The proposed mechanisms underlying the generation of deviance response are described separately in change detection (Chapter 4), regularity formation (Chapter 4), and deviance detection (Chapter 4).

2. **The hypothesis was implemented with biologically plausible models.** Based on the proposed principle, I built simple networks with neural mass models and ran simulations that qualitatively mimic various experimental observations (in simulations I to VI). The neural mass models are parsimonious and have biologically meaningful parameters, which saves simulation time and makes the network dynamics and parameter settings interpretable. The implementation facilitates future extension to other studies in perception.

3. **Model predictions were provided.** The network model based on the proposed principle provides several testable predictions for experimental validation. Some predictions are opposite to what has been proposed by other hypotheses, so future experiments that pinpoint to these conflicting predictions could push forward our knowledge of perception. The model predictions are summarized in Chapter 7.

Lastly, several existing theories of cortical computations in perception were reviewed in Chapter 7. Some theories attempt to answer what task the brain is trying to solve and how the brain solves it. Some theories attempt to approach the problem from a more biological implementation using artificial neural networks. I compared these theories with the proposed principle in an attempt to bridge the gap between the top-down and bottom-up approaches. This would be beneficial for reaching a unifying theory of perception.

1.3   Structure of the thesis

In Chapter 2, the properties of deviance-related responses are described. The existing hypotheses of deviance response, including the proposed generic deviance detection hypothesis, are introduced and compared. The confusion and conflicts among the hypotheses are presented and are discussed in the end.

In Chapter 3, the simulation approach is described. This includes the mathematical descriptions of the neural mass models, the node, the network, the short-term plasticity rules, and the generation of MEG signals.

In Chapters 4 to 6, the two functional stages *change detection* and *regularity formation* are described separately. The *deviance detection* is described with a focus on the interaction between the two functional stages. Simulations I to VI are given to support the proposed hypothesis.

In Chapter 7, the proposed neural mechanisms underlying the generation of deviance response are summarized. The extension of the generic deviance detection principle toward a unifying framework of perception is discussed. Model predictions and future directions are provided in the end.

Chapter 2

Background

"It is not the answer that enlightens, but the question."

– Eugène Ionesco

2.1 Overview

Automatic detection of sudden changes crucially enables reorientation of attention towards relevant events in the environment and thereby is important for survival. From a functional perspective, sensitivity to stimulus deviation likely plays many roles in the nervous system. These may include noise rejection, duration tuning, chunking and grouping, beat perception, and even language comprehension (see reviews in [Näätänen et al., 2007, Xu et al., 2014]). Generally speaking, temporal changes enrich the hierarchical representations of percepts.

The ability to detect abrupt temporal changes is thought to be a pervasive property of the sensory systems, given that deviance-related responses have been widely observed from cellular to system levels, across species, sensory modalities, and spanning from the lower levels of the sensory pathway to the cortex. The sudden change in features at different levels of abstraction elicit various brain responses at different sensory stages. These deviance-related responses, such as the On/Off responses, omitted-stimulus response, mismatch response, left anterior negativity (LAN; syntax violation), and N400 (semantic violation [Bornkessel-Schlesewsky and Schlesewsky, 2019]), have different latencies and sources but may share at

least one common property: They are elicited by a deviant event that violates the perceived regularity[1].

A deviance-related response reflects how the corresponding regularity is perceived. Therefore, the deviance-related responses serve as a probe for investigating neural representation of sensory (or abstract) features, since recording neural representation of features is way more difficult. As illustrated in Figure 2.1, the acoustic stimuli lead to neural representation of a certain regularity (e.g., constant intensity, periodicity, and object identity in the stimuli), and the deviance-related responses (e.g., On/Off responses, OSR, and MMN) reflect the change of the neural representation.
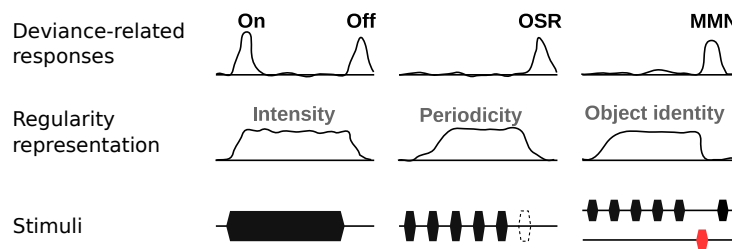
Figure 2.1: The relationship between deviance-related responses and regularity representation. The On/Off responses, omitted-stimulus response (OSR), and mismatch negativity (MMN) in the top row are elicited by the acoustic stimuli in the bottom row (black block: standard tone; dashed block: omission; red block: deviant tone). The neural representations of regular features in the middle row are relatively difficult to capture by the recording/imaging techniques because they may be encoded in a distributed manner. Thus, a unifying theory depicting the relationship between deviance-related responses and regularity representation can be helpful in guiding the direction of investigation.

Compared to a deviance response, the actual neural representation of regularity can be more difficult to capture by the recording/imaging techniques because it can be encoded in a distributed manner in spiking patterns and synaptic connection patterns. Therefore, a theory that depicts the relationship between deviance response and regularity representation can be very helpful for guiding experiment designs.

Below, I review the experimental observations of the deviance-related responses, specifically on the properties of On/Off responses, OSR, and MMN (Section 2.2). I then describe and compare the two main-stream hypotheses of deviance response,

---

[1]The LAN corresponds to a violation to regular phrase structure rules. The N400 corresponds to a violation to regularly meaningful stimuli.

namely the *prediction* and *adaptation* hypotheses, as well as the proposed generic deviance detection principle (Section 2.3).

2.2 Deviance-related responses

2.2.1 On/Off response

The On/Off response is the response of a neuron that shows more spikes at either the onsets, the offsets, or both of a prolonged stimulus. The On/Off responses have been observed at early stages of the auditory pathway. They were observed using extracellular recording in the superior paraolivary nucleus (SPN or SPON) of rodents [Behrend et al., 2002, Dehmel et al., 2002, Felix et al., 2011, Kopp-Scheinpflug et al., 2011, Kulesza Jr et al., 2003], inferior colliculus (IC) of chinchillas [Guo and Burkard, 2002], and the medial geniculate body (MGB) of the guinea pig [He, 2003].

Cortical On/Off responses have been observed using different recording and imaging techniques, including single-cell recording in primary auditory cortex (A1) of awake cats [Chimoto et al., 2002, Qin et al., 2007], and anesthetized rats [Scholl et al., 2010], extracellular recording in A1 of awake marmoset monkeys [Saha et al., 2017], surface micro-electrode array in auditory cortex (AC) of rats [Takahashi et al., 2004], multi-unit extracellular recordings across broad range of AC of mice [Joachimsthaler et al., 2014], flavoprotein fluorescence imaging [Baba et al., 2016] and two-photon calcium imaging [Baba et al., 2016, Deneux et al., 2016] in AC of mice, and MEG in human auditory evoked responses [Nishihara et al., 2014].

The On/Off responses show various waveforms (when presented as firing rate). The studies on the generation of On/Off responses mainly targeted at the peripheral stages (see reviews in [Xu et al., 2014, Kopp-Scheinpflug et al., 2018]). The On responses are shaped by the *adaptation mechanism* and the *post-onset inhibition*. The Off responses are generated as *post-inhibitory rebounds* and firstly arise only when inhibitory connections come into play (Figure 2.2). At later stages, the On/Off responses can be generated by a mixture of excitatory and inhibitory signals, which are simply relayed signals from the preceding stages. However, at the cortical stage new On/Off responses may also be generated through the abundant

reciprocal connections. In this thesis, I have focused on the mechanism of the locally generated On/Off responses at the cortical stage (i.e., change detection in Chapter 4).
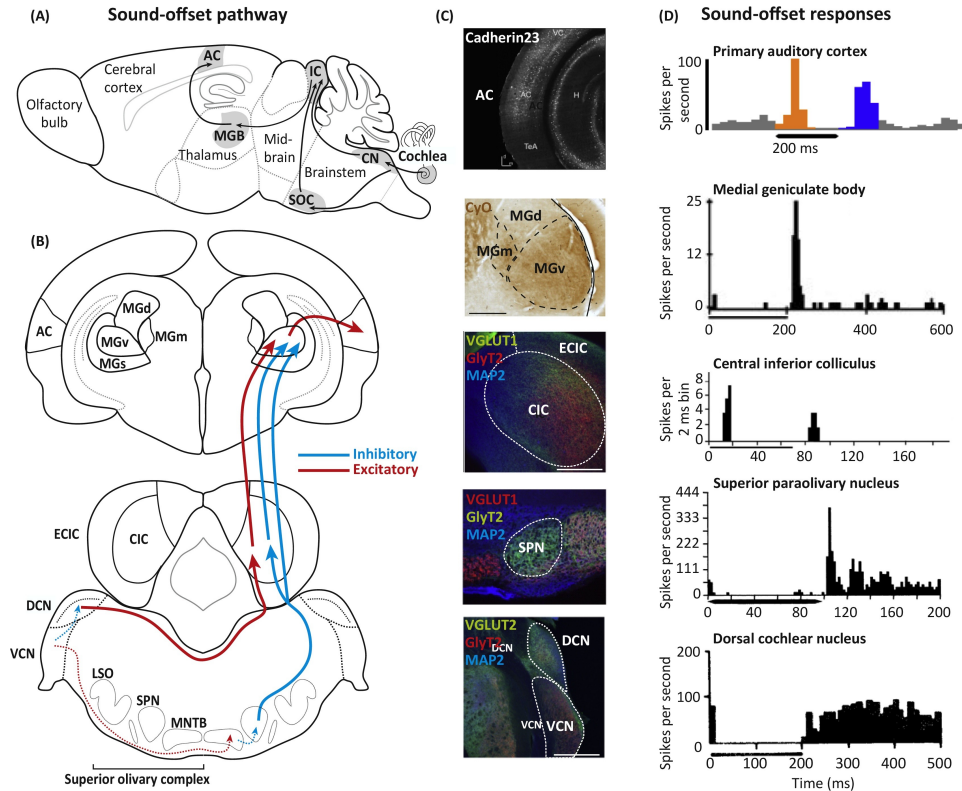
Figure 2.2: Off responses along the auditory pathway. (A) Sagittal schematic of mouse brain. Signals from cochlear go to cochlear nucleus (CN), superior olivary complex (SOC), inferior colliculus (IC), medial geniculate body (MGB), and to the auditory cortex (AC). (B) Coronal schematic. The Off responses are firstly generated in dorsal cochlear nucleus (DCN) and superior paraolivary nucleus (SPN) that receives the inhibitory input (blue dashed line). The Off responses are then relayed and generated at IC, MGB, and AC thought excitatory (red lines) and inhibitory (blue lines) pathways. (C) Coronal sections from mouse brain. (D) Peristimulus time histograms (PSTH) of neural responses in different regions. In AC, neurons often produce both an On response (orange) and an Off response (blue). For more details, please refer to [Kopp-Scheinpflug et al., 2018]. (Figure copied from [Kopp-Scheinpflug et al., 2018])

2.2.2 Omitted-stimulus response

The omitted-stimulus responses/potentials (OSRs/OSPs) is elicited by an omitted stimulus in a periodic stimulus train. OSRs have been observed in different sensory

systems (e.g., visual, auditory, somatosensory) in various species, for example, the visual pathway of fish, reptile, and invertebrate *in vivo* [Bullock et al., 1990, Prechtl and Bullock, 1994, Ramón et al., 2001, Karamürsel and Bullock, 1994], retinas of salamander *in vitro* [Schwartz et al., 2007, Werner et al., 2008], and the electrosensory system of rays [Bullock et al., 1993].

The OSR resembles the Off response as they both peak at the end of a stimulus (or a train of stimuli). However, the peak latency of OSR is time-locked not to the last stimulus but to the omitted one (Figure 2.3), which differentiates the OSR from the Off response. The elicitation of OSR reflects that the nearby neural circuits possess the ability to predict the timing of the next stimulus (i.e., temporal expectancy).

In human EEG/MEG studies, OSRs were observed at the cortical level (often termed omission response, or omission MMN) [Horváth et al., 2010, Andreou et al., 2015, Bullock et al., 1994, Karamürsel and Bullock, 2000, Busse and Woldorff, 2003], but not in the midbrain (IC, tectum) [Nishihara et al., 2014] or the brainstem [Lehmann et al., 2016], where only Off responses are observed. This may reflect that the periodicity is not yet encoded at the stage of IC in humans.
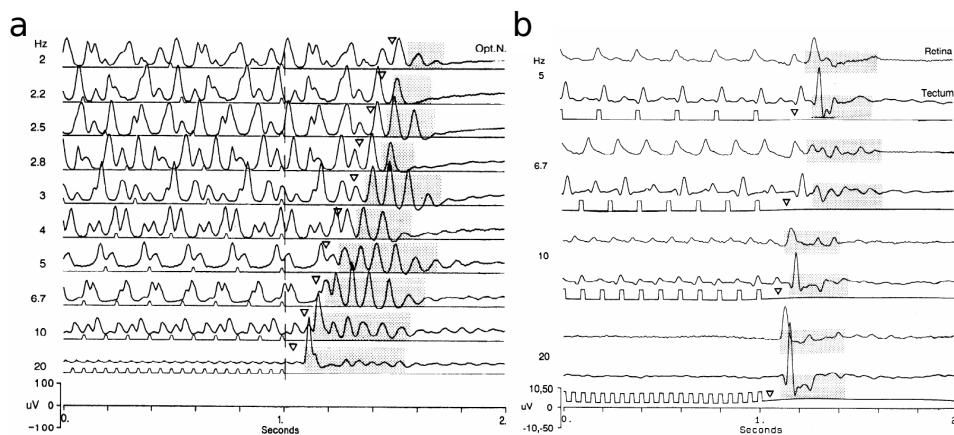
Figure 2.3: The visual-evoked potentials (VEPs) and the omitted-stimulus potentials (OSPs). (a) Local field potentials (LFPs) were recorded in the optic nerve in elasmobranchs and teleosts. The VEPs were evoked by light flashes (periodically in 2Hz to 20 Hz). Similar waveforms as the VEPs can be observed at the missing stimuli (the triangles). The OSPs were elicited after the triangles (the grey areas). (b) The VEPs and OSPs in retina and optic tectum. (Figure copied from [Bullock et al., 1990])

2.2.3   Mismatch response

The mismatch response is elicited by a deviant among repetitive standard stimuli. The underlying process leads to the reorientation of attention to higher cognitive processes. In human audition, the mismatch response is found as a negative deflection (called mismatch negativity, or MMN) in the event-related potential (ERP) with the sources most prominently localized in the auditory cortex.

MMNs have been shown for auditory deviants involving pitch [Sams et al., 1985, Tiitinen et al., 1994, Näätänen et al., 1997, Tervaniemi et al., 2000, Yago et al., 2001a, Yago et al., 2001b, Jacobsen et al., 2003, Novitski et al., 2004, Novitski et al., 2007], intensity [Näätänen et al., 1989, Rinne et al., 2006], duration [Amenedo and Escera, 2000, Ruusuvirta et al., 2013, Näätänen et al., 2004, Hsiao et al., 2010, Näätänen et al., 1989, van Wassenhove and Lecoutre, 2015, Schönwiesner et al., 2007, Jacobsen and Schröger, 2003, Hsu et al., 2010, Colin et al., 2009, Jaramillo et al., 2000], stimulus onset asynchrony (SOA) [Kujala et al., 2001, Tse and Penney, 2006, Brannon et al., 2004], sequence (or pattern) [Schröger et al., 1994, Herholz et al., 2009, Boh et al., 2011, Yaron et al., 2012, Kuchenbuch et al., 2013, Toufan et al., 2016], and more complex features such as rising and falling tones (reviewed in [Paavilainen, 2013]) or voice [Knösche et al., 2002]. A larger deviant magnitude (i.e., larger difference from the standard stimuli) elicits a larger peak amplitude and a shorter peak latency in MMN (Figure 2.4).

The mismatch responses take place at many stages, including in frequency following responses (FFR), middle latency responses (MLR), and long latency responses (LLR) [Shiga et al., 2015]. At early stages, the mismatch responses may be primarily due to the adaptation effect. However, it is also believed that top-down prediction is evolved in the generation of mismatch response even at the stage of IC and MGB [Parras et al., 2017].
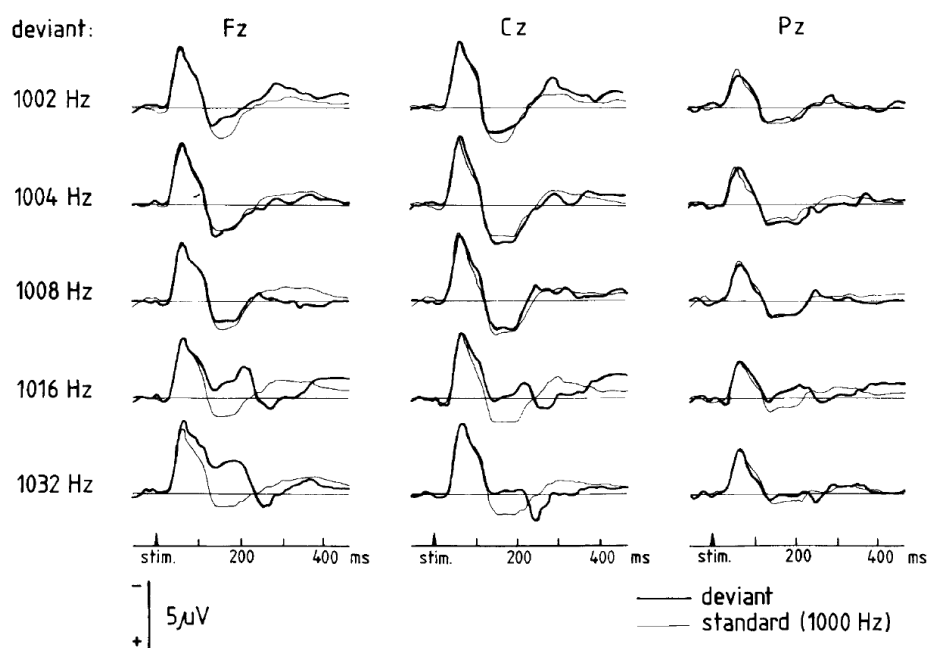
Figure 2.4: The pitch mismatch negativity (MMN) can be found in the more negative curve in the event-related potentials (ERPs) by the deviants (thick line) than by the standards (thin line) at around 200 ms after the onset of the deviant. The MMNs are more prominent at electrodes in frontal (Fz) and vertex (Cz) than in parietal (Pz). (Figure copied from [Sams et al., 1985])

2.3   Hypotheses of deviance response

The term *deviance response* generally refers to the *mismatch response*, and the term *deviance detection* refers to the automatic detection of deviants that elicit mismatch responses in the auditory cortex[2]. The mismatch response reflects the fact that the brain remembers the preceding features (in the standard stimuli) and notices a deviant stimulus that carries different features. It is therefore interesting to know where the memory trace is stored and whether there is a specialized circuit that compares the sensory input and the memory trace.

There are two mainstream hypotheses accounting for the generation of mismatch response. The *prediction hypothesis* (Section 2.3.2) assumes explicit memory traces and specialized circuits as a comparator, while the *adaptation hypothesis*

---

[2]In this thesis, deviance response is a general term covering the deviance-related responses. Its scope may be differently defined elsewhere. The prediction and adaptation hypotheses primarily focus on accounting for the mismatch response.

(Section 2.3.3) assumes implicit encoding of the past stimuli in the reduced synaptic efficacy. The two hypotheses are fundamentally different, but both have support from experimental evidence. I propose the *generic deviance detection hypothesis* (Section 2.3.4) in an attempt to reconcile the confusion and conflict between the two hypotheses.

2.3.1   Prediction hypothesis

The prediction hypothesis emphasizes the need for top-down prediction during perception. Here I choose a neuronal model based on predictive-coding theory [Wacongne et al., 2012] for comparison because this model is more biologically motivated, which makes it easier to compare with other models. The model implements the predictive-coding theory in a straightforward way. For each sensory feature to predict (e.g., tones A and B), there is a Prediction Error layer, a Predictive layer, and a Memory trace unit (Figure 2.5a). The Predictive layer generates top-down prediction signals, and the Prediction Error layer generates bottom-up error signals (see simulations in Figure 2.5b).

The Predictive layer generates prediction signals through the plastic synapses from the Memory trace unit. The Memory trace unit was implemented in a simplified way (i.e., delay lines), which makes the whole model somewhat less biological.
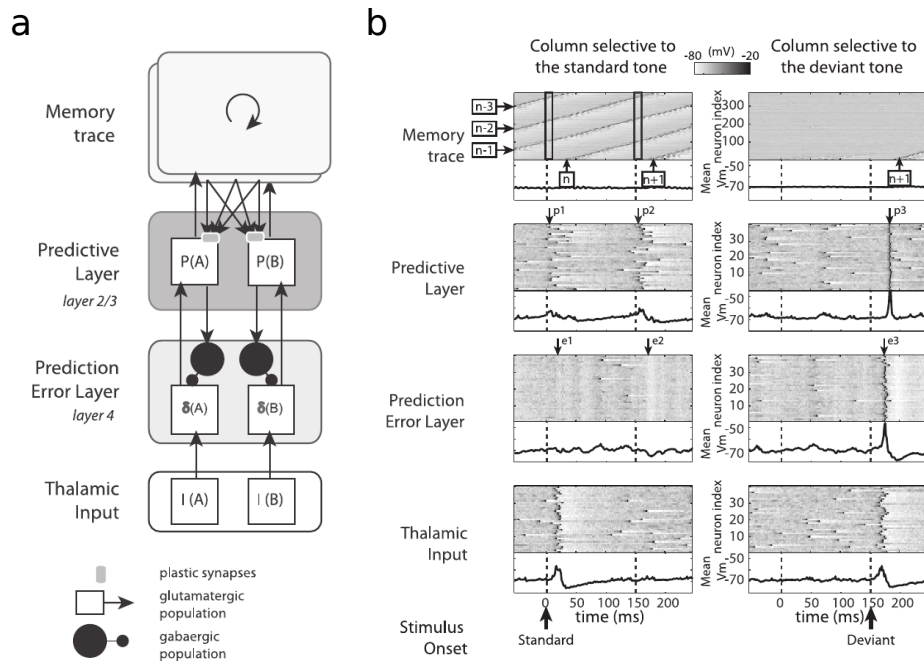
Figure 2.5: A neuronal model of predictive-coding accounting for the MMN. (a) Network structure that implements predictive-coding. The thalamic inputs (tones A and B) arrive at the Prediction Error layer and are inhibited by the prediction signals from the Predictive layer. The prediction error signals then go up to the Predictive layer and the Memory trace unit. The Memory trace unit (implemented in delay lines) keeps the timing of error events and generates prediction signals through the plastic synapses between it and the Predictive layer. The plastic synapses associate the memory signals with the error signals that enter the Memory trace unit. (b) Simulated pattern of neural firing and membrane voltage. The standard stimulus (tone A) arrives at the left column at $t = 0$ ms, and the deviant (tone B) arrives at the right column at $t = 150$ ms. The Prediction Error layer at the left column does not show strong neural activities because the Predictive layer induces hyperpolarization at approximate $t = 0$ and 150 ms. On the other hand, tone B results in prediction error signal (at around $t = 150$ ms), which propagates and updates the memory trace unit at the right column. For more details, please refer to [Wacongne et al., 2012]. (Figure copied from [Wacongne et al., 2012])

2.3.2 Adaptation hypothesis

The adaptation hypothesis emphasizes the role of synaptic adaptation in the generation of mismatch responses. I use an adaptation model by Patrick May [May and Tiitinen, 2010] for illustration because the model has been used to account for several types of mismatch responses[3]. The model has a serial core-belt-parabelt

---

[3]The model settings in [May and Tiitinen, 2010] are different from the settings in [May and Tiitinen, 2013, May et al., 2015]. However, the concepts are similar.

network that comprises four areas (core, belt, parabelt, and P2). Each area consists of 120 sharply-tuned tonotopic columns and 1 broadly-tuned column. The sharply-tuned columns are affected by lateral inhibition from neighboring columns. All columns have excitatory connections to themselves and others. Synaptic adaptation is applied to the excitatory connections. Thalamic inputs only go to the core area.

The simulation result in Figure 2.6 shows that in this model the MMN is simply a delayed N1[4], where the delayed activation is mainly due to the adaptation and lateral inhibition in the sharply-tuned columns. This suggests that there is no need for specialized circuits to achieve deviance detection.

---

[4]N1 is an ERP that peaks between 80 to 120 ms after the onset of a stimulus.
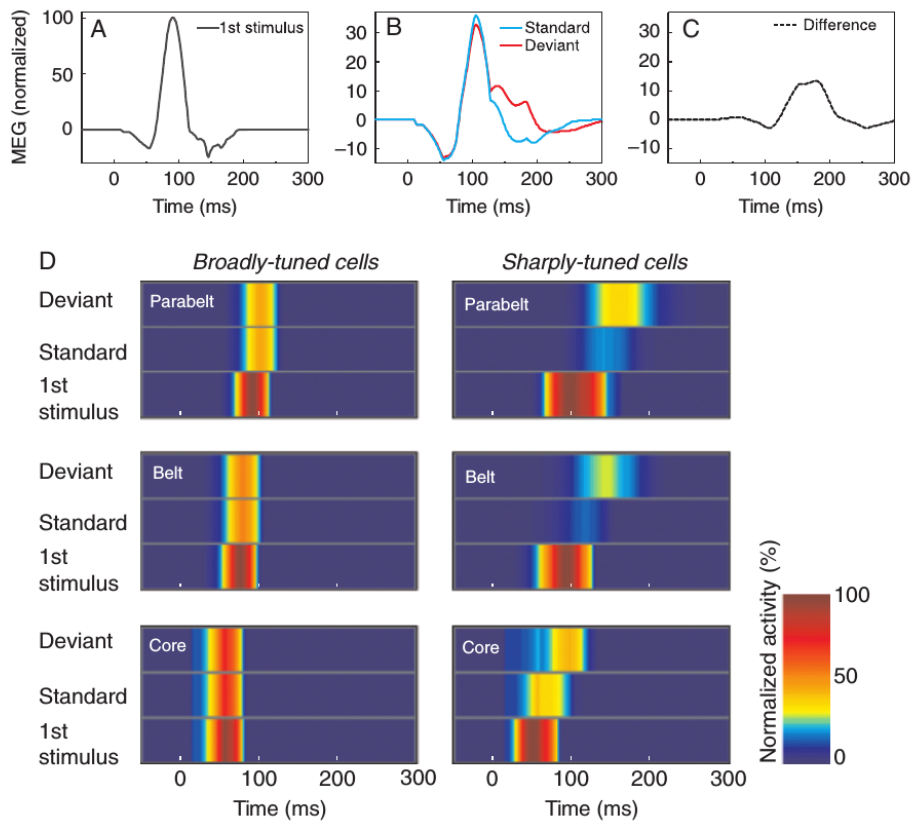
Figure 2.6: Simulations of the adaptation model demonstrating that MMN may be simply delayed N1 response. The model has a core-belt-parabelt network structure, with each area comprising sharply tuned and broadly tuned populations (For detailed structure, see the Appendix in [May and Tiitinen, 2010]). (A) Simulated MEG signal as N1 response elicited by the 1st standard tone. (B) The standard tones elicit N1 responses (blue). The deviant tone elicits a similar N1 (red) except that the curve at 120-220 ms is stronger. (C) The MMN (difference between the two N1 responses in B) peaks around $t = 180$ ms. (D) Breakdown of the activities in the core, belt, and parabelt areas. The sharply-tuned cells (especially in parabelt) show delayed activation for deviants than for standards. The longer N1 latency to a deviant stimulus is due to the effects of both adaptation and lateral inhibition. The adaptation model suggests that the MMN is simply a delayed N1 response and that there is no dedicated MMN generator. (Figure copied from [May and Tiitinen, 2010])

2.3.3 Generic deviance detection hypothesis

The proposed generic deviance detection hypothesis rests on the assumption that the process of deviance detection can be functionally separated into stages of *regularity formation* and *change detection*. As illustrated in Figure 2.7a, the hierar-

chical features are constituted by nodes representing various types of feature (f). The R nodes and C nodes can be distributed in the hierarchy, producing features of regularity (f_R) and change (f_C). The R-C pairs serve the function of deviance detection.

The implementation of the generic deviance detection hypothesis does not require specialized neural circuits. As illustrated in Figure 2.7b, the function of change detection in Node 2 emerges under a proper setting of reciprocal connections between Nodes 1 and 2 (see details in Chapter 4). The function of regularity formation can be achieved by accumulating bottom-up features through convergent connections (see details in Chapter 5).

I assume cortical deviance detection is supported by neural circuits of a common structural motif, given the relatively uniform wiring patterns across areas in the cortex. In addition, many of the deviance-related activities, though originating from different stages of the auditory pathway, can be observed pervasively in the auditory cortex (as described in Section 2.2). Taken together, the proposed hypothesis suggests that the cortical deviance-related activities are primarily generated locally through reciprocally connected neural circuits.
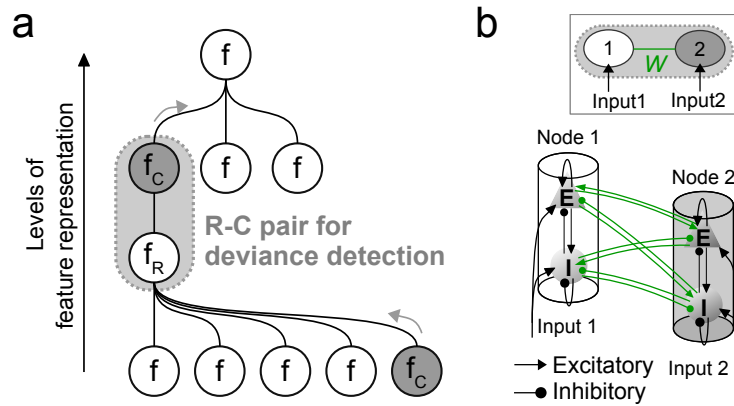
Figure 2.7: Illustration of the role of deviance detection in hierarchical feature representation. (a) The process of feature representation includes the interaction between regularity formation (R) and change detection (C). The R nodes remain stable (regularity, f_R) by accumulating the ascending information from the lower-level features. The C nodes detect abrupt temporal changes in the neighboring R node(s) and pass them to the higher levels as new features (f_C, gray arrows). In this sense, an R-C pair forms a basic mechanism of deviance detection which takes place at every level in the hierarchy. (b) An R-C pair is formed by two reciprocally coupled nodes. In the simulations, all nodes are allowed to receive external weighted inputs that reach the excitatory and inhibitory populations. The inter-node connections (green) are the free parameters, and the intra-node connections are fixed for simplicity.

2.3.4   Comparison of hypotheses

The prediction and adaptation hypotheses described in Section 2.3.1 and 2.3.2 were used specifically to account for the mismatch response. The proposed generic deviance detection hypothesis takes into consideration the On/Off responses and the omitted-stimulus response as well. There are conceptual differences between the three hypotheses. I illustrate the schematics in Figure 2.8 and point out the main differences. Clarifying the difference helps to resolve the conflicts and confusions between the hypotheses.
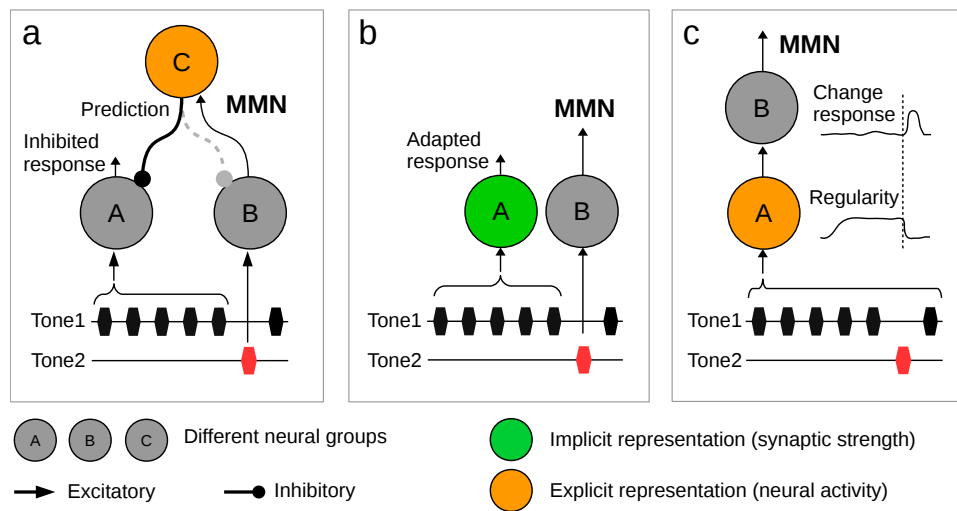
Figure 2.8: Comparison of models of mismatch negativity. (a) Prediction-based model. In an oddball paradigm, the standard stimuli (tone 1, black) reach neural group A, and the deviant stimulus (tone 2, red) reaches neural group B. Group A generates small error signals (inhibited responses) in response to the standard stimuli, because group A is inhibited (black round arrowhead) by prediction from group C (orange circle). Group B generates a more significant error signal (MMN) in response to the deviant stimulus because group B is not inhibited (grey round arrowhead). The error signals are used to update group C (the internal model) for future prediction. (b) Adaptation-based model. Group A (green circle) responds less to the standard stimuli due to the adaptation effect. Group B shows a stronger response (MMN) to the deviant (red) because the synapses in group B is not adapted. (c) The generic-deviance-detection model. Group A (orange circle) shows sustained activity (regularity) after a few repetitions of standard stimuli. The sustained activity falls due to either a deviant or an omission. Group B generates a change response (MMN) in response to the change in group A activity.

First, the hypotheses suggest different mechanisms underlying the generation of MMN. The prediction hypothesis suggests that the mismatch response is an error signal between the prediction and the actual sensory input. The adaptation hypothesis suggests that the mismatch response is a delayed N1. The Generic deviance detection hypothesis suggests the mismatch response is a change response to regularity. In terms of MMN source, both the prediction hypothesis and adaptation hypothesis imply that the MMN source lies at (or close to) the neural populations selective to the deviant tone (group B in Figure 2.8 a and b). The generic deviance detection hypothesis implies that the MMN source lies at the neural populations that is sensitive to the change in regularity (group B in Figure 2.8c), which is not necessarily selective to the deviant tone.

Second, the hypotheses imply different ways to encode regularity. In the prediction hypothesis, regularity (or prediction) is represented explicitly in the firing pattern of higher level populations (orange circle in Figure 2.8a; can be in a distributed manner). In the adaptation hypothesis, regularity is encoded implicitly in the synaptic strengths (green circle in Figure 2.8b). In the generic deviance detection hypothesis, regularity is represented in lower level populations (orange circle in Figure 2.8c; can be in a distributed manner). Both the adaptation hypothesis and the generic deviance detection hypothesis imply that regularity is locally encoded in the neural populations that directly receive sensory inputs. Conversely, the prediction hypothesis implies that the sensory inputs do not directly form the regularity.

2.4 Confusion and conflict to be resolved

I have compared the conceptual differences among the hypotheses of deviance detection (Section 2.3.4). Below I raise some issues regarding the confusion and conflict among the hypotheses. These issues underscore the need for a unifying view of deviance detection, covering the cortical On/Off responses, the cortical OSR, and the MMN.

1. *Which neural circuits give rise to the diverse cortical On/Off responses?*
   Knowledge of the generation of On/Off responses has been mainly derived from observations at non-cortical stages. The On responses are thought to be due to adaptive and post-onset inhibitory mechanisms that shape the responses in the auditory nerve [Phillips et al., 2002]. The Off responses are widely accepted to arise from post-inhibitory rebound (see review in [Kopp-Scheinpflug et al., 2018] for the detailed cellular and synaptic mechanisms), as concluded from observation in SPON neurons [Felix et al., 2011]. Other response patterns such as On-Off and On-sustained-Off[5] can then potentially be explained by mixing of excitatory and inhibitory inputs with different delays in a feed-forward network [Xu et al., 2014]. As for the On/Off responses recorded in the auditory cortex, they may originate from the ascending non-cortical On/Off responses [Scholl et al., 2010] or be generated locally in the cortex.

---

[5]The On-Off and On-sustained-Off patterns in [Xu et al., 2014] correspond to the Dec-OnOff and Inc-OnOff types mentioned in Section 4.2.2, respectively.

Not much attention has been focused on how On/Off responses are generated locally in the cortex, since the above mechanisms have explained various On/Off waveforms. However, this does not mean the cortex does not generates On/Off responses locally. The cortical On/Off neurons show diverse temporal profiles [Deneux et al., 2016]. Also, a single cortical neuron may have distinct onset- and offset-frequency receptive fields (FRFs) [Qin et al., 2007]. Do these properties of cortical On/Off responses rise from the feedforward combinations from the thalamus or recurrent wiring in the cortex?

2. *Is the omitted-stimulus response (OSR) just sustained resonance?*
   The OSR has two properties. First, the peak latency includes an additional constant delay (e.g., around 100 ms in human MEG/EEG) from the time when the missing stimulus would have occurred (due-time). It does not depend on the stimulus-onset asynchrony (SOA) [Andreou et al., 2015, Schwartz and Berry II, 2008]. Second, the peak amplitude can be larger than the entrained responses during periodic stimuli [Horváth et al., 2010].

   Some modeling studies suggested that the OSR is sustained resonance (i.e., damping oscillations) [May and Tiitinen, 2001, Thivierge and Cisek, 2011]. This may be wrong, because sustained resonance alone cannot explain the additional delay and higher peak amplitude in OSR. How the neural circuits maintain the input periodicity and detect the change is unclear.

3. *Does the OSR reflect prediction or prediction error?*
   This question rests on whether the OSR is triggered by a similar mechanism as the MMN.The MMN is thought to reflect either a prediction-error signal resulting from the comparison between the input and the top-down prediction (prediction hypothesis), or an increased signal caused by the stimulus propagating through un-adapted synapses (adaptation hypothesis).

   According to the computational models based on either hypothesis, the OSR is qualitatively different from the classical MMNs elicited by other deviants. The adaptation-based model suggests the OSR to be a rebound response (i.e., sustained resonance) rather than a modulated N1 [May and Tiitinen, 2001]. The prediction-based model suggests the OSR to reflect predictive signals rather than prediction error [Wacongne et al., 2012]. Both interpretations implicitly suggest the OSR to be purely endogenous activities, which con-

flict with the two properties of ORS mentioned above (i.e., the additional constant delay and the larger peak amplitude). The mechanism underlying the generation of OSR is therefore not yet convincingly explained by the existing models.

Chapter 3

Methods

"Yin and Yang (the two opposing forces) is so-called Dao (the principle)."

– I Ching (The Book of Changes)

3.1 Neural mass model

Neural mass models are rate-based models that allow the construction of neural networks to be simple and scalable. In addition, the simulation results based on the neural mass model can be compared to the experimental observations such as LFP and MEG/EEG. Therefore, neural mass models are good candidates for study in the network dynamics that are to be linked with behavioral and physiological evidence.

To model deviance-related responses, I create simple recurrent networks, where a node in the network represents a certain cortical area in the auditory cortex (Figure 3.1). Each node comprises two neural populations. The mathematical descriptions of *neural population*, *node*, and *network* are given in Sections 3.2, 3.3, and 3.4. I used different types of *short-term plasticity* in the simulations, including synaptic adaptation and function-driven plasticity rules. These rules are described in Section 3.5. I also generated *simulated MEG signals* from the model to compare the simulation results with the experimental observations. The equations for generating MEG signals are described in Section 3.6. Finally, the default settings are listed in Section 3.7.
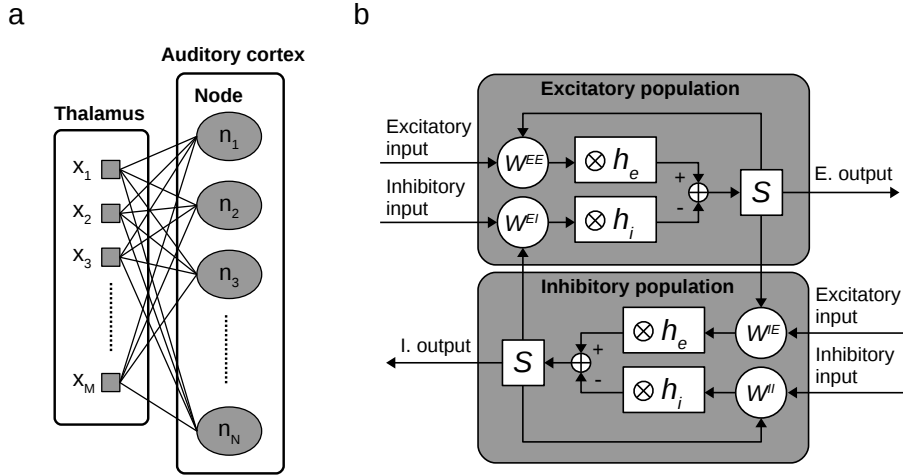
Figure 3.1: The network model for simulation. (a) The auditory cortex is represented by a recurrent network that consists of N all-to-all connected nodes (connections not shown). The nodes receive M thalamic inputs $x(t)$ through external connections (W^{EX} and W^{IX}). For nodes that represent Belt and Parabelt areas, the external connection weights are set to zero. (b) A node consists of one excitatory (E) and one inhibitory (I) neural population. In each population, there are two rate-to-potential operators ($\otimes h_e$ and $\otimes h_i$) and a potential-to-rate operator (S). The intra- and inter-node connection strengths are specified by W^{EE}, W^{IE}, W^{EI}, and W^{II} (external connections W^{EX} and W^{IX} not shown).

3.2 Neural population

The processing of neural activities in a neural population can be separated into two stages. As illustrated in Figure 3.2, the first stage covers how the presynaptic spikes affect the postsynaptic potential (PSP), and the second stage covers the generation of spikes at the axon hillocks. The two stages are governed by the *rate-to-potential operator* (RPO) and the *potential-to-rate operator* (PRO), respectively.

For the first stage, the RPO describes a linear transformation from input mean firing rate $x(t)$ to mean PSP $v(t)$. The inputs $x_c(t)$, where $c \in \{e,i\}$, reach the excitatory/inhibitory synapses and are transformed to the EPSP/IPSP $v_c(t)$. The transformation mimics the dynamics of the synapses and dendrites, and is achieved by convolving the input $x_c(t)$ with a synaptic kernel $h_c(t)$.

$$v_c(t) = x_c(t) \otimes h_c(t) \tag{3.1}$$

The $v_c(t)$, including the EPSP $v_e(t)$ and the IPSP $v_i(t)$, contribute to the overall PSP $v(t)$.

$$v(t) = v_e(t) - v_i(t) \tag{3.2}$$

For the second stage, the PRO transforms the overall PSP $v(t)$ into the output firing rate $m(t)$ by a non-linear sigmoid function $S(\cdot)$ as described in Equation 3.3. This mimics the output nonlinearity at the axon hillocks.

$$m(t) = S\big(v(t)\big) \tag{3.3}$$
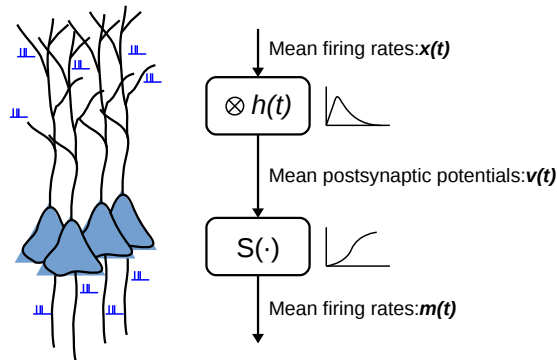
Figure 3.2: The two operators in a neural population. Spikes (blue color) arriving at the dendrites cause the changes in postsynaptic potentials, and the neurons generate spikes if the membrane potential reaches the threshold for depolarization. The process can be separated into two operations. The rate-to-potential operator (denoted by $\otimes h(t)$) transform mean firing rate $x(t)$ to mean postsynaptic potential $v(t)$. The potential-to-rate operator (denoted by $S(\cdot)$) transforms $v(t)$ to meaning firing rate $m(t)$ by a sigmoid function. See text in detail.

3.2.1 Jansen-Rit-based operators

The two operators used in the Jansen-Rit model [Jansen and Rit, 1995, Jansen et al., 1993, Spiegler et al., 2010, Spiegler et al., 2011] are more biological, and the values of the operator parameter were set according to experimental evidence. I used the Jansen-Rit-based operators for the simulations in Chapters 4 and 6.

For the PRO, the sigmoid function $S(\cdot)$ is described in Equation 3.4, where e_0 controls the maximum firing rate, and r controls the slope at the membrane potential v_0 for firing.

$$S(v) = \frac{2e_0}{1 + e^{r(v_0 - v)}} \tag{3.4}$$

For the RPO, the two synaptic kernels $h_c(t)$, where $c \in \{e, i\}$, are described in Equation 3.5. The average synaptic gain H_c controls the peak value of the response curve. The time constant τ_c represents the delay due to dendritic effects and neurotransmitter kinetics.

$$h_c(t) = \begin{cases} \frac{H_c}{\tau_c} t e^{\frac{-t}{\tau_c}} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{3.5}$$

The convolution of the input $x_c(t)$ with the kernel $h_c(t)$ can be further represented by a second-order differential equation. This was derived using the Laplace transformation $\mathcal{L}\{\}$ as below.

$$\mathcal{L}\Big\{v_c(t)\Big\} = \mathcal{L}\Big\{x_c(t) \otimes h_c(t)\Big\} \tag{3.6}$$

$$V_c(s) = X_c(s) \left(\frac{H_c}{\tau_c} \frac{1}{(s + \frac{1}{\tau_c})^2} \right) \tag{3.7}$$

$$\left(s^2 + \frac{2s}{\tau_c} + \frac{1}{\tau_c^2} \right) V_c(s) = \frac{H_c}{\tau_c} X_c(s) \tag{3.8}$$

$$\mathcal{L}^{-1}\left\{ \left(s^2 + \frac{2s}{\tau_c} + \frac{1}{\tau_c^2} \right) V_c(s) \right\} = \mathcal{L}^{-1}\left\{ \frac{H_c}{\tau_c} X_c(s) \right\} \tag{3.9}$$

$$\left(\frac{\partial^2}{\partial t^2} + \frac{2}{\tau_c} \frac{\partial}{\partial t} + \frac{1}{\tau_c^2} \right) v_c(t) = \frac{H_c}{\tau_c} x_c(t) \tag{3.10}$$

In the numerical simulation, I implemented the RPO in Equation 3.10 by two first-order ordinary differential equations (Equations 3.11 and 3.12)

$$\dot{v}_c(t) = u_c(t) \tag{3.11}$$

$$\dot{u}_c(t) = \frac{H_c}{\tau_c} x_c(t) - \frac{2}{\tau_c} u_c(t) - \frac{1}{\tau_c^2} v_c(t) \tag{3.12}$$

3.2.2 Wilson-Cowan-based operators

The two operators used in the Wilson-Cowan model [Wilson and Cowan, 1972, May and Tiitinen, 2013, May et al., 2015] are relatively simplified. I used the Wilson-Cowan-based operators for the simulations in Chapter 5 because there are more nodes involved in the network, and the simulations using Wilson-Cowan-based operators can be faster than using Jansen-Rit-based operators.

For the PRO, the sigmoid function $S(\cdot)$ is described in Equation 3.13, where r controls the slope and v_0 is the threshold for firing. $[x]_+$ denotes $max(x,0)$.

$$S(v) = \left[\tanh\left(r(v - v_0) \right) \right]_+ \tag{3.13}$$

For the RPO, the two synaptic kernels $h_c(t)$, where $c \in \{e, i\}$, are described in Equation 3.14.

$$h_c(t) = \begin{cases} \frac{1}{\tau_c} e^{\frac{-t}{\tau_c}} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{3.14}$$

The convolution of the input $x_c(t)$ with the kernel $h_c(t)$ can be represented by a first-order differential equation. The derivation using Laplace transformation is as below.

$$\mathcal{L}\left\{ v_c(t) \right\} = \mathcal{L}\left\{ x_c(t) \otimes h_c(t) \right\} \tag{3.15}$$

$$V_c(s) = X_c(s) \left(\frac{1}{\tau_c} \frac{1}{s + \frac{1}{\tau_c}} \right) \tag{3.16}$$

$$\left(s + \frac{1}{\tau_c} \right) V_c(s) = \frac{1}{\tau_c} X_c(s) \tag{3.17}$$

$$\mathcal{L}^{-1}\left\{\left(s+\frac{1}{\tau_c}\right)V_c(s)\right\} = \mathcal{L}^{-1}\left\{\frac{1}{\tau_c}X_c(s)\right\} \tag{3.18}$$

$$\left(\frac{\partial}{\partial t}+\frac{1}{\tau_c}\right)v_c(t) = \frac{1}{\tau_c}x_c(t) \tag{3.19}$$

In the numerical simulation, I implemented the RPO in Equation 3.19 as below.

$$\dot{v}_c(t) = \frac{-1}{\tau_c}v_c(t) + \frac{1}{\tau_c}x_c(t) \tag{3.20}$$

3.3   Node

The node represents a basic unit in a hierarchical feature representation shown in Figure 2.7. Each node consists of one excitatory (E) and one inhibitory (I) neural population. Depending on the applications, such a two-population node can be treated as (a) partial/one/many cortical column(s). Here I treat a node as a *functional unit* rather than a structurally-specific one.

The state variables of the E and I population in node k comprise the PSPs $v_k^p(t)$ and the output firing rate $m_k^p(t)$, where the superscript $p \in \{E,I\}$ stands for the excitatory/inhibitory population. The PSPs $v_k^p(t)$ are computed from the EPSPs $v_{k,e}^p(t)$ and the IPSPs $v_{k,i}^p(t)$ as below.

$$v_k^p(t) = v_{k,e}^p(t) - v_{k,i}^p(t) \tag{3.21}$$

The output firing rates $m_k^p(t)$ are calculated as below.

$$m_k^p(t) = S\left(v_k^p(t)\right) \tag{3.22}$$

Neural populations interact with each other by means of firing rate via the synaptic connections specified by the matrices W^{EE}, W^{IE}, W^{EI} and W^{II} which correspond to E-to-E, E-to-I, I-to-E, and I-to-I connections, respectively. Self-feedback is allowed. See the network descriptions in Section 3.4.

3.4   Network

For N nodes that represent N locations in the auditory cortex and M external inputs that represent the intensity of M tones (as shown in Figure 3.1), the network structure is defined by four $N \times N$ connection matrices W^{EE}, W^{IE}, W^{EI} and W^{II}, and two $N \times M$ external connection matrices W^{EX} and W^{IX}. Each element w (non-negative) in the connection matrices stands for the gain factor on the firing rate, which reflects the number and strengths of the synapses established from the source to the target population. The element w_{kj}^{EI} of W^{EI}, for example, stands for the connection strength from the I population in node j to the E population in node k. The E populations are fed with constant background input B. The external stimuli $x(t)$ reach the E and I populations via external connections W^{EX} and W^{IX}.

If the Jansen-Rit-based operators (mentioned in Section 3.2.1) are to be used in the simulation, the dynamics of $v_{k,e}^E(t)$, $v_{k,i}^E(t)$, $v_{k,e}^I(t)$, and $v_{k,i}^I(t)$ are described as below. The input $x_q(t)$ stands for the intensity of tone q, and the background input B_k is constant for node k. The short-term plasticity terms a_{kj}, p_{kj}^{EE}, p_{kj}^{IE}, p_{kj}^{EI}, and p_{kj}^{II} modulate the connection strengths w_{kj}^{EE}, w_{kj}^{IE}, w_{kj}^{EI}, and w_{kj}^{II}, respectively (see more details on short-term plasticity in Section 3.5).

$$\left(\frac{\partial^2}{\partial t^2} + \frac{2}{\tau_e}\frac{\partial}{\partial t} + \frac{1}{\tau_e^2} \right) v_{k,e}^E(t) = \frac{H_e}{\tau_e} \left[\sum_{j=1}^{N} a_{kj} p_{kj}^{EE} w_{kj}^{EE} m_j^E(t) + \sum_{q=1}^{M} w_{kq}^{EX} x_q(t) + B_k \right]$$

$$(3.23)$$

$$\left(\frac{\partial^2}{\partial t^2} + \frac{2}{\tau_i}\frac{\partial}{\partial t} + \frac{1}{\tau_i^2} \right) v_{k,i}^E(t) = \frac{H_i}{\tau_i} \left[\sum_{j=1}^{N} p_{kj}^{EI} w_{kj}^{EI} m_j^I(t) \right] \qquad (3.24)$$

$$\left(\frac{\partial^2}{\partial t^2} + \frac{2}{\tau_e}\frac{\partial}{\partial t} + \frac{1}{\tau_e^2} \right) v_{k,e}^I(t) = \frac{H_e}{\tau_e} \left[\sum_{j=1}^{N} p_{kj}^{IE} w_{kj}^{IE} m_j^E(t) + \sum_{q=1}^{M} w_{kq}^{IX} x_q(t) \right] \qquad (3.25)$$

$$\left(\frac{\partial^2}{\partial t^2} + \frac{2}{\tau_i}\frac{\partial}{\partial t} + \frac{1}{\tau_i^2} \right) v_{k,i}^I(t) = \frac{H_i}{\tau_i} \left[\sum_{j=1}^{N} p_{kj}^{II} w_{kj}^{II} m_j^I(t) \right] \qquad (3.26)$$

Similarly, if the Wilson-Cowan-based operators (mentioned in Section 3.2.2) are to be used in the simulation, the dynamics of $v_{k,e}^E(t)$, $v_{k,i}^E(t)$, $v_{k,e}^I(t)$, and $v_{k,i}^I(t)$ are described as below.

$$\left(\frac{\partial}{\partial t} + \frac{1}{\tau_e}\right)v_{k,e}^E(t) = \frac{1}{\tau_e}\left[\sum_{j=1}^{N} a_{kj}p_{kj}^{EE}w_{kj}^{EE}m_j^E(t) + \sum_{q=1}^{M} w_{kq}^{EX}x_q(t) + B_k\right] \quad (3.27)$$

$$\left(\frac{\partial}{\partial t} + \frac{1}{\tau_i}\right)v_{k,i}^E(t) = \frac{1}{\tau_i}\left[\sum_{j=1}^{N} p_{kj}^{EI}w_{kj}^{EI}m_j^I(t)\right] \quad (3.28)$$

$$\left(\frac{\partial}{\partial t} + \frac{1}{\tau_e}\right)v_{k,e}^I(t) = \frac{1}{\tau_e}\left[\sum_{j=1}^{N} p_{kj}^{IE}w_{kj}^{IE}m_j^E(t) + \sum_{q=1}^{M} w_{kq}^{IX}x_q(t)\right] \quad (3.29)$$

$$\left(\frac{\partial}{\partial t} + \frac{1}{\tau_i}\right)v_{k,i}^I(t) = \frac{1}{\tau_i}\left[\sum_{j=1}^{N} p_{kj}^{II}w_{kj}^{II}m_j^I(t)\right] \quad (3.30)$$

3.5   Short-term plasticity

There are various forms of short-term plasticity, which includes short-term depression/facilitation (lasting for hundreds of milliseconds to seconds), augmentation (lasting for tens of seconds), and post-tetanic potentiation (lasting for minutes). Short-term plasticity is thought to underlie information processing. I considered short-term depression (synaptic adaptation, Section 3.5.1) and other function-driven rules (Section 3.5.2) in the simulations.

3.5.1   Short-term depression (synaptic adaptation)

Synaptic adaptation is a ubiquitous phenomenon. In this thesis, the effect of synaptic adaptation was considered in most simulations[1]. The synaptic adaptation term

---

[1]The effect of synaptic adaptation was not considered only in some simulations in Chapter 4 for the purposes of demonstration and comparison.

a_{kj} ($j, k \in \{1, 2, ..., N\}$) in Equations 3.23 and 3.27 modulates the efficacy of connection w_{kj}^{EE}. The value of a_{kj} varies only according to the pre-synaptic activity $m_j^E(t)$.

$$\dot{a}_{kj}(t) = \frac{1 - a_{kj}(t)}{\tau_a} - \kappa_a a_{kj}(t) m_j^E(t) \qquad (3.31)$$

The time constant τ_a represents the recovery rate of the synaptic efficacy, and the constant κ_a influences the decay rate of $a_{kj}(t)$. The value of $a_{kj}(t)$ ranges from 0 to 1, and is fixed to 1 if synaptic adaptation is not considered in the simulation.

3.5.2 Other function-driven rules

I assume that other forms of short-term plasticity would be also involved in the process of regularity formation[2], for example, using the short-term depression/facilitation that depend both on pre- and post-synaptic activities[3] so that the sequential patterns can be encoded.

In a general form, the plasticity term p_{kj} ($j, k \in \{1, 2, ..., N\}$, and j$\neq$k) is used to modulate the efficacy of connection w_{kj}, with $0 \leq p_{kj} \leq 1$ for short-term depression, and $p_{kj} \geq 1$ for short-term facilitation. The plasticity rule is described in Equation 3.32. The time constant τ_p controls the recovery rate toward the stable point P, and κ_p controls the learning rate. The function $g(\cdot)$ considers the effectiveness of plasticity regarding the geometrical distance between nodes j and k. This can be done by, for example, a Gaussian function $exp\left(-(j-k)^2/2\sigma^2\right)$, where σ controls the mask width of effectiveness. The function $f(\cdot)$ increases/decreases p_{kj} according to the pre-synaptic information $x_j(t)$, the post-synaptic information $x_k(t)$, and the status of $p_{kj}(t)$.

$$\dot{p}_{kj}(t) = \frac{P - p_{kj}(t)}{\tau_p} + \kappa_p \cdot g(j,k) \cdot f\left(x_j(t), x_k(t), p_{kj}(t)\right) \qquad (3.32)$$

---

[2]In other words, these function-driven rules were not used in change detection.

[3]Short-term plasticity are mostly dependent on the pre-synaptic activities. A dependence on post-synaptic activity may exist but the underlying mechanisms are not well studied.

3.6   Simulated MEG signal

I simulate MEG signals from the activity of neural populations for comparison with the experimental observations. Physiologically, the MEG signals originate mainly from the *primary currents*[4] in the spatially aligned pyramidal neurons, where the primary currents include the Ca^{2+} and Na^+ ion currents through excitatory synapses and the Cl^- ion currents through inhibitory synapses [da Silva, 2004, Jackson and Bolger, 2014]. See illustration in Figure 3.3.

(1) Excitatory synapse on apical dendrites   (2) Inhibitory synapse near soma   (3) Excitatory synapse near soma   (4) Inhibitory synapse on apical dendrites
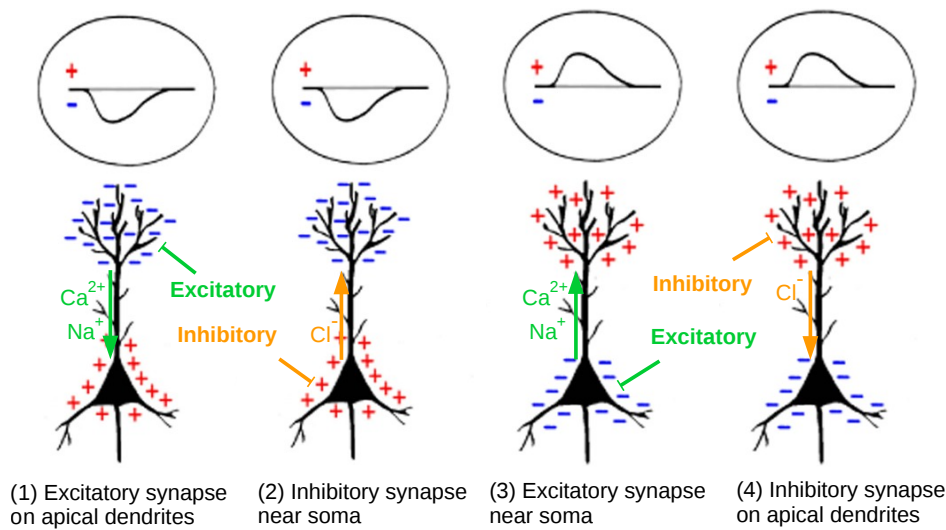
Figure 3.3: Physiological basis of EEG. The first and second conditions cause negative deflections in the EEG, and the third and four conditions cause positive deflections. For a special case when a pyramidal cell receives excitatory inputs on apical dendrites and inhibitory inputs near soma, the cell may not fire but both the positive and negative ion currents contribute to a negative deflection in the EEG. (Images adapted from [Jackson and Bolger, 2014])

I assume that the excitatory synapses are mainly on apical dendrites and the inhibitory synapses are mainly near the soma. The assumption leads to the summed contribution of the ions currents. The simulated MEG signal $R_k(t)$ of node k is therefore calculated according to Equation 3.33, where both $m^E(t)$ and $m^I(t)$ contribute to the MEG amplitude, with $r_{current1}$ and $r_{current2}$ specifying the portions.

---

[4]The primary currents are intracellular currents in the post-synaptic neurons, and the secondary currents are are extracelluar currents that close the current loop.

$$R_k(t) = \sum_{j=1}^{N} \left[r_{current1} \cdot a_{kj} w_{kj}^{EE} m_j^E(t) + r_{current2} \cdot w_{kj}^{EI} m_j^I(t) \right] \qquad (3.33)$$

The overall MEG signal $R(t)$ is the weighted sum of $R_k(t)$ as below, where $\sum_k^N r_k = 1$. The purpose of applying weights r_k is only to highlight the activities of specific nodes (e.g., the change detectors). By default the weights are set equally.

$$R(t) = \sum_{k=1}^{N} r_k \cdot R_k(t) \qquad (3.34)$$

3.7   Setting of parameters

In Chapters 4 and 6, demonstrations of change detection and deviance detection are run in networks using the Jansen-Rit-based operators. The general configurations are listed in Table 3.1. The parameter settings of neural population model are kept the same as proposed by Jansen and Rit [Jansen and Rit, 1995, Jansen et al., 1993], unless specified otherwise. The intra-node connections were fixed in order to reduce the number of free parameters. The values of intra-node connections are chosen such that a single node stays inactivated under weak excitatory input and starts to oscillate as the excitatory input strength increases to e_0 (i.e., half of the maximum value of the sigmoid function). The value of constant background input B_k was the same across nodes and was chosen such that the nodes work in proper conditions (i.e., near a bifurcation point for an isolated node). The external inputs $x_{(t)}$ reaches both the excitatory and inhibitory populations. The adaptation parameters τ_a and κ_a are chosen such that a single node remains oscillating during prolonged stimulation, rather than showing only a transient peak response at the onset.

In Chapter 5, demonstrations of regularity formation are run in networks using the Wilson-Cowan-based operators. The general configurations are listed in Table 3.2. The parameter settings were adjusted from the settings in [May and Tiitinen, 2010, May and Tiitinen, 2013, May et al., 2015]. The exact settings were specified in the respective Sections.

Table 3.1: General configurations (Jansen-Rit-based operators)

Par.	Value	Unit	Description
Impulse response function			
τ_e	10	ms	Time constant (excitatory synapse)
τ_i	20	ms	Time constant (inhibitory synapse)
H_e	3.25	mV	Gain (excitatory synapse)
H_i	22	mV	Gain (inhibitory synapse)
Sigmoid function			
e_0	2.5	spikes/s	Maximum firing rate
r	0.56	1/mV	Slop
v_0	6	mV	Threshold for firing
Intra-node connections			
w_{kk}^{EE}	135×0.8	1	E-to-E connection in node k
w_{kk}^{IE}	135×0.6	1	E-to-I connection in node k
w_{kk}^{EI}	135×0.2	1	I-to-E connection in node k
w_{kk}^{II}	135×0.05	1	I-to-I connection in node k
Inter-node connections			
w_{kj}^{EE}	$135 \times \{0, 0.1, ..., 0.5\}$	1	E_j-to-E_k connection
w_{kj}^{IE}	$135 \times \{0, 0.1, ..., 0.5\}$	1	E_j-to-I_k connection
w_{kj}^{EI}	$135 \times \{0, 0.1, 0.2\}$	1	I_j-to-E_k connection
w_{kj}^{II}	$135 \times \{0, 0.1, 0.2\}$	1	I_j-to-I_k connection
External connections			
w_{kq}^{EX}	220×0.2	1	x_q-to-E_k connection
w_{kq}^{IX}	220×0.1	1	x_q-to-I_k connection
Background inputs and external inputs			
B_k	220×0.5	spikes/s	Background input to E_k
$x_q(t)$	-	spikes/s	External input q (step function, amplitude= 1.5, rise/fall time= 10ms)
Synaptic adaptation			
τ_a	200	ms	Recovery rate
κ_a	2	1	Drop rate

Table 3.2: General configurations (Wilson-Cowan-based operators)

Par.	Value	Unit	Description
Impulse response function			
τ_e	20	ms	Time constant (excitatory synapse)
τ_i	60,80,100	ms	Time constant (inhibitory synapse)
Sigmoid function			
r	2/3	1/mV	Slop
v_0	0	mV	Threshold for firing
Intra-node connections			
w_{kk}^{EE}	6	1	E-to-E connection in node k
w_{kk}^{IE}	0.5 to 5	1	E-to-I connection in node k
w_{kk}^{EI}	0.5 to 5	1	I-to-E connection in node k
w_{kk}^{II}	0	1	I-to-I connection in node k
Inter-node connections			
w_{kj}^{EE}	-	1	E_j-to-E_k connection (short-term plastic)
w_{kj}^{IE}	-	1	E_j-to-I_k connection (short-term plastic)
w_{kj}^{EI}	0	1	I_j-to-E_k connection
w_{kj}^{II}	0	1	I_j-to-I_k connection
External connections			
w_{kq}^{EX}	1	1	x_q-to-E_k connection
w_{kq}^{IX}	0	1	x_q-to-I_k connection
Background inputs and external inputs			
B_k	-	spikes/s	Background input to E_k
$x_q(t)$	-	spikes/s	External input q (step function, amplitude= 1, rise/fall time= 10ms)
Synaptic adaptation			
τ_a	200,600	ms	Recovery rate
κ_a	2,20	1	Drop rate

Chapter 4

Change detection

"What defines us is how well we rise after we fall."

– Zig Ziglar

4.1 Overview

Change detection is a function that detects abrupt temporal changes in a signal. The On response and Off response naturally serve this function, as they respond to the edges of a signal. Supposing that the cortical wiring pattern is naturally suitable for the emergence of the cortical On/Off responses, change detection can be ubiquitously involved in many higher cognitive functions[1].

In Section 4.2, I use two reciprocally-connected nodes in the simulations to reproduce the properties of the cortical On/Off responses, a demonstration to support the notion of ubiquitous local change detection. In Section 4.3, I examine the underlying connection patterns that generate the simulated On and Off responses. I then investigate how altered connection patterns (e.g., reduced external connections to inhibitory populations, effect of NMDA-r antagonists, and synaptic adaptation) affect the emergence of change detectors. At the end of this chapter, I summarize the key messages suggested by the simulation results.

---

[1] A *change detector* is different from a *comparator* that checks the difference between two signals. The pursuit of a generic wiring pattern for change detection differs from the concept of predictive coding theory that suggests the need for the comparison between prediction signals and sensory signals. Change detection compares the integral of previous signal with the current signal.

4.2   Cortical On/Off responses

The cortical On/Off responses have several properties. These include (1) diverse temporal profiles across neurons and (2) distinct onset and offset frequency receptive fields (FRFs) in a neuron. A follow-up question is: *Can cortical circuits give rise to the observed properties of cortical On/Off responses?*

In Section 4.2.1, I review some experimental observations on the two properties. Based on the assumption that the cortex also generates On/Off responses locally (i.e., not just inherited from the non-cortical On/Off responses) and that the cortical wiring is relatively uniform, I demonstrate that a simple network with reciprocal connections can account for the two observed properties (Sections 4.2.2, and 4.2.3). The simulation results suggest that change detection can take place in a wide range of cortical areas.

4.2.1   Experimental observations

The cortical On/Off responses are diverse and complex. A prolonged tone stimulus can elicit diverse temporal patterns of On/Off responses in the auditory cortex. Neurons can be sensitive to the onset/offset of the stimulus (i.e., transient responses at the edges) and also show increased or decreased firing rate during the stimulus (i.e., level changes) compared with the spontaneous activity [Chimoto et al., 2002, Qin et al., 2007, Joachimsthaler et al., 2014, Deneux et al., 2016, Recanzone, 2000, Volkov and Galazjuk, 1991]. The temporal profiles depend on the properties of neurons (e.g., locations) and the properties of stimuli (e.g., pitch, intensity, duration, etc.). I use two experimental observations that show high non-linear properties for demonstration. The observed properties are to be matched with those of the simulations.

First, cortical neurons show diverse temporal profiles in response to the stimuli. As shown in Figure 4.1a, the labeled neurons formed spatially intermingling clusters, where the labeling was based on clustering of the responses to the eight stimuli (Figure 4.1b). The homogeneity index of each label (representing an average fraction of neighboring cells within a 30 μm radius that belonged to the same label) was computed to quantify the distribution of cluster. The homogeneity (of 12 out of 13 labels) was significantly higher than in a shuffled map (Figure 4.1c), which suggested spatial clustering of the diverse On/Off responses. In simulation I (Sec-

tion 4.2.2), I defined eight types of On/Off responses and showed how the spatial clustering might be due to the connection patterns in the network.
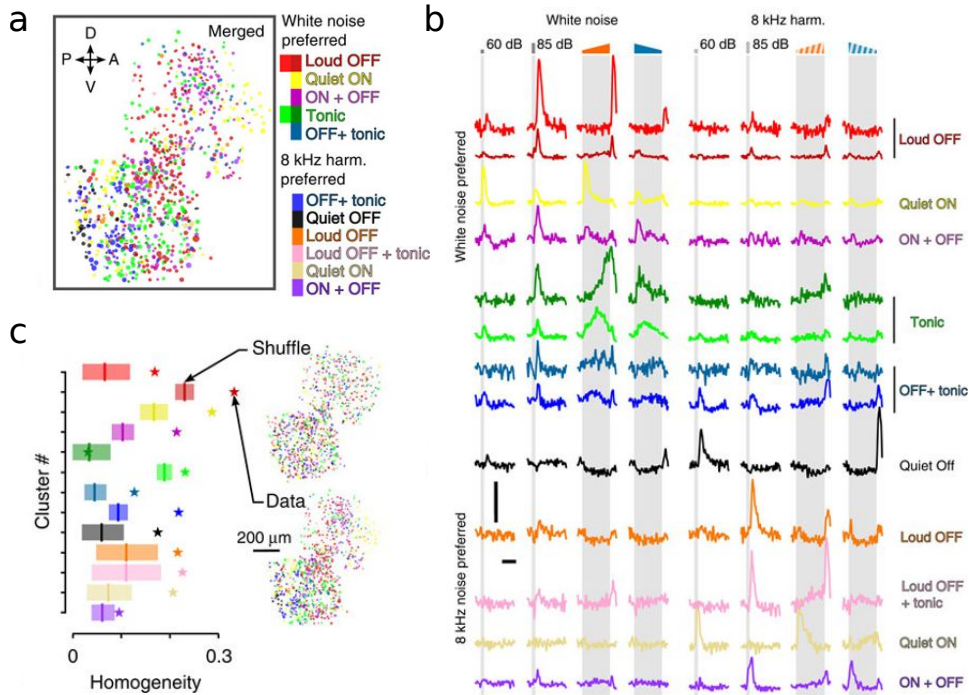

Figure 4.1: Clusters of cortical On/Off responses in awake mice (Two-photon calcium imaging). (a) Localizations of the cells belonging to the different labels (color coded on the right) in an example mouse. (b) Mean temporal profiles of the identified clusters in response to {white noise, 8 kHz tone} × {60 and 85 dB (0.25s duration), up- and down-ramps (2s duration)}. (c) Homogeneity index (color stars) for each label (or 'cluster' in the original paper [Deneux et al., 2016])) calculated across totally three mice. The vertical lines represent the value of Homogeneity obtained by shuffling. (Figure copied from [Deneux et al., 2016])

Second, an individual cortical neuron shows distinct onset- and offset-FRFs. As shown in Figure 4.2, the responses to tonal stimuli depend not only on neuron location but also on tone frequency. For example, a neuron that shows phasic-tonic response to the stimulus of its best frequency (BF) may show Off responses to the stimuli of other frequencies (Figure 4.2h). The On responses and Off responses were not firmly linked together, and the authors suspected that the On/Off responses are generated by independent mechanisms [Joachimsthaler et al., 2014]. In simulation II (Section 4.2.3), I showed how the distinct onset- and offset-FRFs can be explained in a network level.
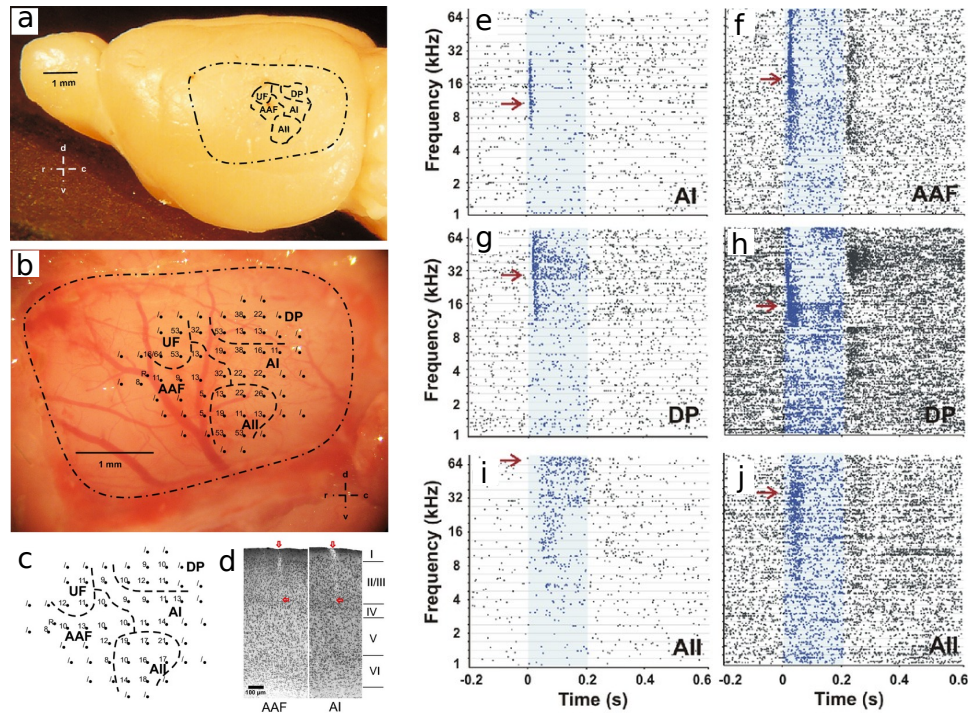
Figure 4.2: Distinct onset- and offset-FRFs in awake mice (extracellular recordings). (a) The electrodes were located in cortical layer III/IV in the primary auditory field (AI), anterior auditory field (AAF), higher-order fields (AII), dorsoposterior field (DP) and ultrasonic field (UF). (b) Enlarged view of the auditory cortex of the same sample mouse, where the number shows the best frequency (BF) at the indicated location. The slashes stand for no tone-evoked responses. (c) The number shows the shortest tone-evoked latency (ms) at the indicated location. (d) The electrode tracks ended in the lower layer III (marked by the red arrows). (e-j) Examples of tone-evoked responses in dot plots of spiking activity. The red arrows indicate the BFs. Duration: 2000 ms (blue areas). Intensity: 70 dB SPL. Tone frequency: 1-64 kHz. (Figure copied from [Joachimsthaler et al., 2014])

4.2.2 Simulation I: diverse On/Off responses

Simulation settings

The input stimulus (duration= 2000 ms, amplitude= 1.5 spikes/s, rise/fall time= 10 ms) was fed to a two-node network (Figure 4.3a). The settings followed the general configurations for networks using Jansen-Rit-based operators specified in Table 3.1. Only node 1 received the input stimulus, and node 2 (the change detector) was connected with node 1 (i.e., the external connections to node 2, $w_2^{EX} = w_2^{IX} = 0$).

Varying the inter-node connections W alters the response $m_2^E(t)$ of the change detector. I scanned a range of inter-node connections ($W^{EE}, W^{IE} \in \{0, 0.1, ..., 0.5\}$; $W^{EI}, W^{II} \in \{0, 0.1, 0.2\}$) and categorized each of the time courses of $m_2^E(t)$ as one of the eight types based on the level changes and the peak at edges: (1) Inc-None, (2) Inc-On, (3) Inc-Off, (4) Inc-OnOff, (5) Dec-None, (6) Dec-On, (7) Dec-Off, (8) Dec-OnOff, and (9) others. (See Figure 4.3b and Table 4.1 for details of categorization.) The 'Inc' and 'Dec' stand for increased and decreased activity during the stimulus. The 'On', 'Off', and 'OnOff' stand for transient peak(s) only at the onset, the offset, or both, of the stimulus. 'None' stands for no clear peaks at the edges of the stimulus. Bistable or non-responsive behaviors were categorized as others.

Simulation results

The *W solutions* are connection settings that give rise to one of the eight categorized On/Off types under this specific simulation settings (e.g., the intensity and onset/offset time of stimulus, the intensity of background input, and intra-node connections, etc). To further investigate the relation between the inter-node connections W and the On/Off responses, I projected the W solutions $\{W_{type\_i}, i = 1, 2, ..., 8\}$ to a 2D plane by t-Distributed Stochastic Neighbor Embedding [Maaten and Hinton, 2008]. This allowed the visualization of the mutual proximity of W solutions in the original eight-dimensional space. The projected W solutions in Figure 4.3d showed several properties. (1) Although the W solutions exhibit a clustered pattern from a broad perspective, different types are observed at small mutual distance when zooming in. The clustering patterns and their sensitivity to W may potentially explain the diverse but spatially clustered On/Off responses shown in Figure 5 of [Deneux et al., 2016]. (2) The Off types are not constrained within Inc/Dec clusters, suggesting that Off responses are not crucially determined by the level change of $m_2^E(t)$ during the stimulus. (3) The On and Off types occupy distinct areas in the 2D plane, which agrees with the conclusion that On and Off responses are driven by largely nonoverlapping sets of synaptic inputs [Scholl et al., 2010]. (4) However, there are also areas where the On, Off and OnOff types are close to each other, where neuroplasticity (e.g., synaptic adaptation, spike-timing-dependent plasticity, or homeostatic plasticity) may play a role in changing the neural response from one type to another.
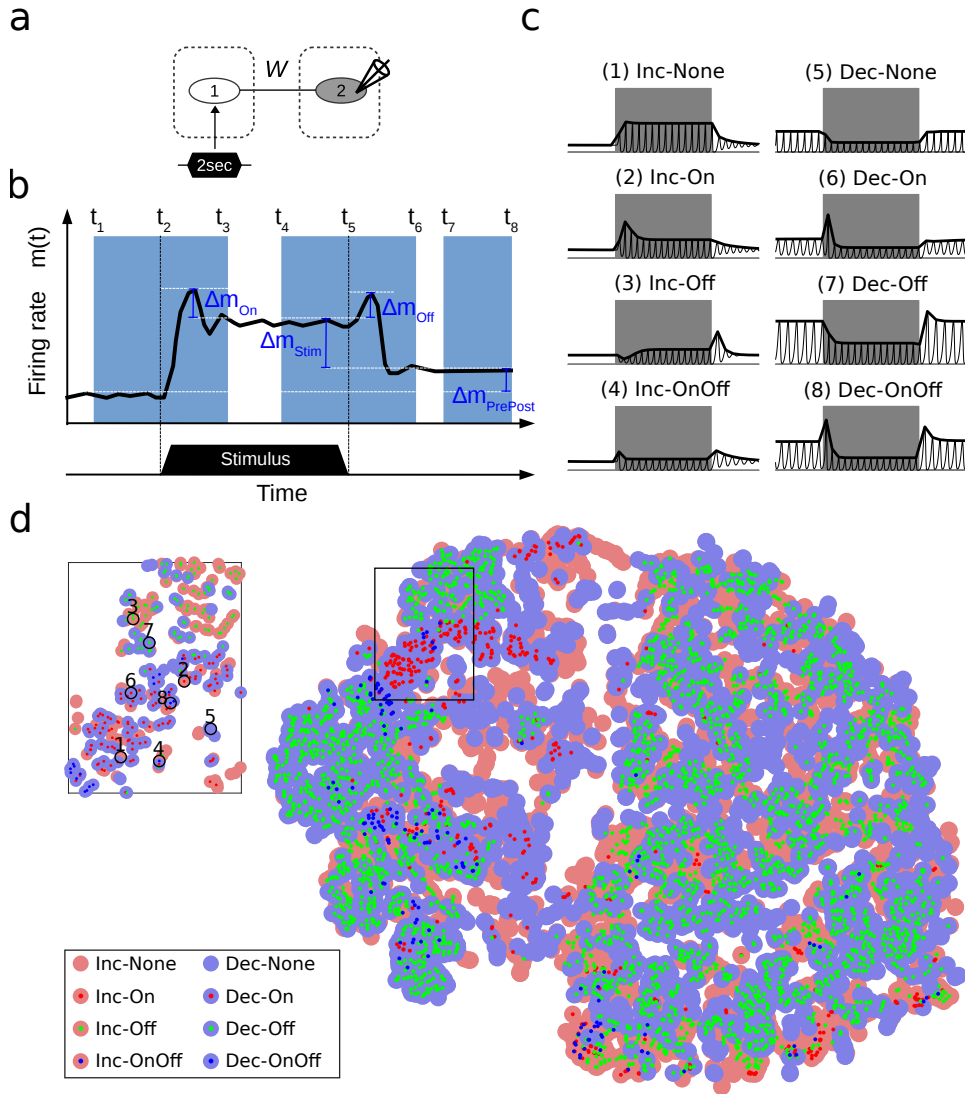
a

b

c

(1) Inc-None   (5) Dec-None

(2) Inc-On   (6) Dec-On

(3) Inc-Off   (7) Dec-Off

(4) Inc-OnOff   (8) Dec-OnOff

d

● Inc-None   ● Dec-None
● Inc-On   ● Dec-On
● Inc-Off   ● Dec-Off
● Inc-OnOff   ● Dec-OnOff

Figure 4.3: Change detectors and the corresponding W solutions. (a) In the simulation settings a prolonged stimulus of 2000 ms is fed to node 1. A range of inter-node connections W are scanned through and the various temporal behaviors of the change detector (i.e., time courses of $m_2^E(t)$) are categorized. (b) For categorization, four variables Δm_{On}, Δm_{Stim}, Δm_{Off}, and $\Delta m_{PrePost}$ are calculated according to the time windows (light blue areas) for each time course $m_2^E(t)$. The time courses that are not bistable are then categorized as one of the eight On/Off types. See detailed categorization settings in Table 4.1. (c) Exemplary responses of eight On/Off types. Gray bands represent the duration of stimulus. The black curves represent the time courses of $m_2^E(t)$, and the bold black curves are the envelopes. (d) All W solutions of the eight On/Off types in the scanned range are projected to a 2D plane for visualization (Matlab function: tsne), where color dots represent the eight types ({Inc, Dec} × {None, On, Off, OnOff}). The eight exemplary behaviors in (c) are labeled in the zoomed in area.

Table 4.1: Settings and variables for categorization of network behavior

Par.	Value	Unit	Description
Time windows			
$[t_{On}, t_{Off}]$	$[0, 2000]$	ms	Stimulus onset and offset
$[t_1, t_2]$	$[-500, 0]$	ms	Pre-onset window
$[t_2, t_3]$	$[0, 500]$	ms	Post-onset window
$[t_4, t_5]$	$[1500, 2000]$	ms	Pre-offset window
$[t_5, t_6]$	$[2000, 2500]$	ms	Post-offset window 1
$[t_7, t_8]$	$[3500, 4000]$	ms	Post-offset window 2
Variables comparing $max\left(m^E(t_i < t < t_j)\right)$			
$\Delta m_{PrePost}$	-	spikes/s	Bistability check ($[t_1, t_2]$ vs. $[t_7, t_8]$)
Δm_{Stim}	-	spikes/s	Level change ($[t_4, t_5]$ vs. $[t_1, t_2]$ and $[t_7, t_8]$)
Δm_{On}	-	spikes/s	Onset peak height ($[t_1, t_2]$ vs. $[t_2, t_3]$)
Δm_{Off}	-	spikes/s	Offset peak height ($[t_4, t_5]$ vs. $[t_5, t_6]$)
Thresholds			
$\theta_{PrePost}$	0.1	spikes/s	**Bistable** behavior, if $\Delta m_{PrePost} \geq \theta_{PrePost}$
θ_{Stim}	0	spikes/s	**Increased** response, if $\Delta m_{Stim} > \theta_{Stim}$
			Decreased response, otherwise
θ_{On}	0.5	spikes/s	**On** response, if $\Delta m_{On} > \theta_{On}$
θ_{Off}	0.5	spikes/s	**Off** response, if $\Delta m_{Off} > \theta_{Off}$

4.2.3 Simulation II: distinct onset- and offset-FRFs

As demonstrated in simulation I, the two-node network can account for the different temporal profiles of On/Off responses. A network with the same properties can account for the distinct onset and offset FRFs in individual cells in the auditory cortex [Qin et al., 2007, Joachimsthaler et al., 2014]. For example, the exemplary cell in Figure 4.4a is sensitive to the onsets of sound stimuli at higher frequencies (3,200-15,872 Hz) and the offsets of sound stimuli at lower frequencies (512-16,000 Hz), as reflected by higher spike density (yellow and red). In addition, this cell shows suppressed spike density (deep blue) during stimuli at low and middle frequencies. In short, the On/Off responses vary across tonal frequencies and across cells.

Simulation settings

The two-node network was used to reproduce the distinct FRFs in Figure 4.4a-c. The simulation input (duration= 500 ms, amplitude= 1.5 spikes/s, rise/fall time= 10 ms) represented a pure tone in one trial of the experimental recordings. The tone frequency was changed for each experimental trial. This was done by altering the parameters $ratio1$ and $ratio2$ in the simulation (Figure 4.4d). The ratios reflected how strongly the two nodes were influenced by the stimulus. Considering the tonotopic organization in the auditory cortex, the ratios were changed for each simulation trial because the stimulus input in each trial represented a different tonal frequency. The simulation input was fed to both nodes with different external connection strength (i.e., $w_1^{EX} = 44 \times ratio1$; $w_1^{IX} = 22 \times ratio1$; $w_2^{EX} = 44 \times ratio2$; $w_2^{IX} = 22 \times ratio2$). The inter-node connections W were picked from the W solutions, were fixed in each example, and the ratios adjusted such that the responses $m_2^E(t)$ of node 2 qualitatively mimicked the experimental observations. The simulation trials were then merged to make simulated FRFs (Figure 4.4e-g).

Simulation results

The simulated FRFs in Figure 4.4e-g reproduced the observed FRFs in Figure 4.4a-c, respectively. In Figure 4.4e the population E_2 showed a Dec-Off response when $ratio1 = 1$ and $ratio2 = 0$ (the same as the ideal case used in simulation I). The On response emerged as $ratio1$ decreased, and a small amount of $ratio2$ resulted

in stronger On responses and weaker Off responses. In Figure 4.4f population E_2 showed a Dec-Off response when $ratio1 = 1$ and $ratio2 = 0$, and turned into Inc-None type when $ratio2$ was larger than $ratio1$. In Figure 4.4g population E_2 showed a Dec-OnOff response, and $ratio2$ enhanced the On responses.

Simulation II provided a sense of how the exemplary cells in Figure 4.4a-c were influenced by different sound tones: $ratio2$ (green side-bar in Figure 4.4e-g) indicated which tones were more directly influencing the recorded cell, whereas $ratio1$ (orange side-bar) reflected how its surrounding neurons were sensitive to the tonal scope. The connections W reflected the interaction between the recorded cell and its surrounding neurons. For example, the side-bar in Figure 4.4f reflects that the recorded cell in Figure 4.4b may be slightly sensitive to the tone at 6.4 to 9.6 kHz, and its neighboring neurons can be strongly sensitive to the tone at around 12.8 kHz.
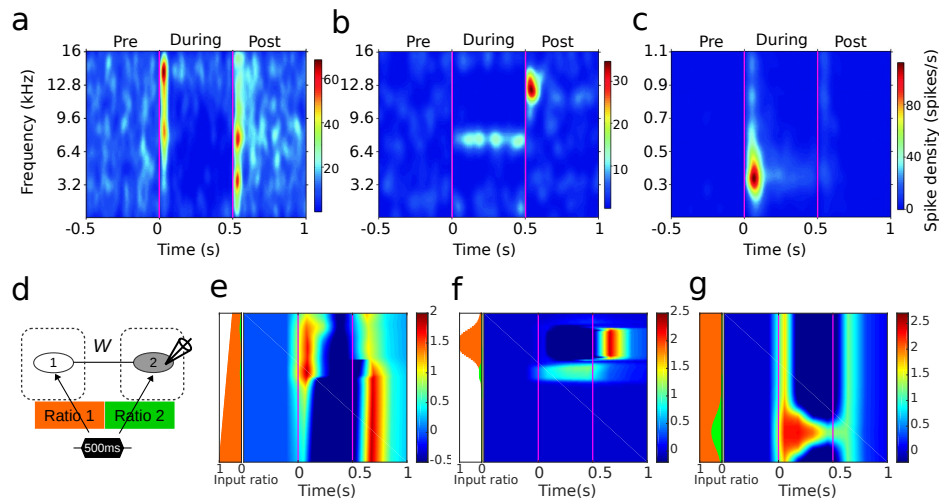

Figure 4.4: Distinct onset and offset FRFs. (a-c) Three exemplary cells that show distinct onset and offset FRFs (Images adapted from [Qin et al., 2007]). The cells were recorded in the primary auditory cortex of awake cats. Pure tones (ranging from 128Hz to 16,000Hz) were presented for 500 ms. The pre-, during-, and post-stimulus spike density of the cell is color coded. (d) In the simulation settings, a two-node network with adjustable external connections w_1^{EX} and w_2^{EX} (orange and green color) are used to mimic the experimental observations. (e-g) Simulation results that mimic the observations in (a-c). The green and orange input ratios at the left-side bar of each plot represent the settings of external connections w_1^{EX} and w_2^{EX} for each simulation trial. The firing rate is color coded. The pre-stimulus firing rate is used as baseline, and the negative value (deep blue color) during the stimulus represents decreased activity.

4.3   Mechanism underlying change detection

In the previous simulation examples, I demonstrated that the behavior of a change detector can account for many phenomena (e.g., diverse cortical On/Off responses, distinct onset- and offset-FRFs). Here, I present a more detailed analysis of the exact requirements for a change detector to work. First, I investigated how and under which conditions the On and Off responses occur. Then, I examined how changes in connection strengths affects the generation of On/Off responses, where the connection strengths were affected by three factors: (1) external input to inhibitory populations, (2) blockage of NMDA receptor channels, and (3) synaptic adaptation.

4.3.1   Onset detection

It has been proposed that On responses could be due to *adaptive* and *post-onset inhibitory* mechanisms that reshape the onset response in auditory nerve fibers [Phillips et al., 2002]. In simulation I, I found that the On responses can also be due to the transiently inhibited activity of population I_2 at the onset of a stimulus. As shown in the magenta rectangles in Figure 4.5a and b, population I_2 is briefly inhibited (indicated by the red arrows) by population I_1, and the transient low $v_2^I(t)$ in turn leads to a transient peak in $v_2^E(t)$ (indicated by the black arrows). The system returns to stability soon after the $v_2^E(t)$ peak brings $v_2^I(t)$ up again. These On responses caused by *transient disinhibition* suggest an important role of the inter-node connection W^{II} in the generation of the On responses.
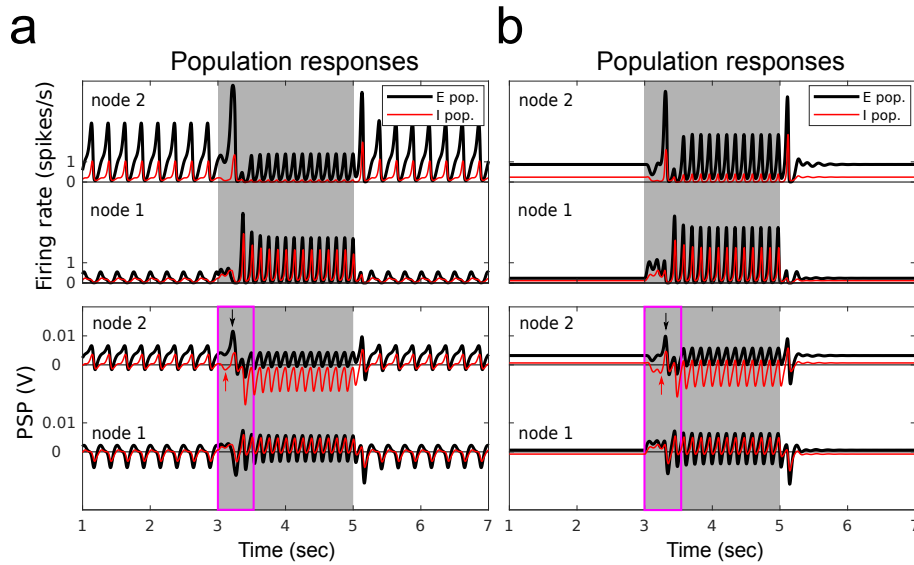
Figure 4.5: Onset detection. (a) In the simulation settings, a prolonged stimulus of 2000 ms is fed to the R node (node 1) in a two-node network. The inter-node connections W is chosen from the W solutions that give rise to Dec-OnOff responses in the C node (node 2) (see Figure 4.3b). The node responses to the stimulus (gray period) are shown as firing rates (upper plot) and as PSPs (lower plot). As an Dec-OnOff response, the time course $m_2^E(t)$ (black curve) shows a lower amplitude during the stimulus and transient peaks at the onset and offset of the stimulus. To understand the generation of On response in population E_2, the magenta rectangle indicates the period of transient disinhibition where population I_2 is transiently inhibited by population I_1 (red arrow), and population E_2 peaks right after the transient disinhibition (black arrow). (b) A similar example for Inc-OnOff response, where the generation of an On response is also due to the transient disinhibition.

4.3.2   Offset detection

It is widely accepted that Off responses that follow decreased activity (i.e., the Dec-Off responses) arise from *post-inhibitory rebound* that is related to the intrinsic conductance property of the neuronal membranes [Kopp-Scheinpflug et al., 2011]. However, the generation of Off responses that follow increased activity (i.e., the Inc-Off responses) cannot be simply explained by the post-inhibition mechanism (see reviews in [Xu et al., 2014, Kopp-Scheinpflug et al., 2018]). Here, I investigate under which conditions the Dec-Off and Inc-Off responses might arise at the network level.

The time courses of two typical Dec-Off and Inc-Off responses are shown in Figure 4.6a and c. In the upper panels, the population E_2 showed Off responses (i.e., $m_2^E(t)$ in black curves), regardless of its decreased/increased activity during the stimulus

(gray area). The lower panels show that both Dec-Off and Inc-Off responses can result from the same mechanism: *a release from long-lasting disinhibition*. The Off response came in two steps. First, the population I_2 received strong inhibition from population I_1 during the stimulus (reflected by the negative PSP $v_2^I(t)$, red curves in the green rectangles). Second, the population E_2 activity peaked before I_2 recovered after stimulus offset (the transient peak $v_2^E(t)$, black curves in the magenta rectangles).

The generation of Off responses can also be illustrated by the phase portraits shown in Figure 4.6b and d. The trajectories of the phase portraits showed how $v_2^E(t)$ and $v_2^I(t)$ evolved interactively. When there was only background input, E_2 and I_2 oscillated in the normal steady state (the blue counter-clockwise trajectories) where E_2 excites I_2, and I_2 inhibits E_2. During stimulus presentation, I_1 became active, and E_2 and I_2 oscillated in a 'reversed' steady state (the green clockwise trajectories) because E_2 had an additional inhibitory effect on I_2 through the pathway $E_2 \rightarrow I_1 \rightarrow I_2$, and I_2 had an additional disinhibitory effect on E_2 through the pathway $I_2 \rightarrow I_1 \rightarrow E_2$. The Off responses were depicted by the magenta trajectories during the transition from the reversed steady state to the normal steady state.

Simulation II provides clues for the underlying neural mechanisms. The inter-node connection W^{II} is critical for a network to give rise to the Off responses because the inhibitory population I_2 first has to be inhibited (i.e., disinhibition). The inter-node connection W^{EI} is important to maintain the network in the working state (e.g., the reversed steady state) otherwise the network gets 'overheated' during disinhibition. With these structural prerequisites, the excitatory population E_2 may show a transient Off response before the inhibited population I_2 catches up again following stimulus offset.

The timing of stimulus offset (i.e., the initial point in the state space when the transition begins) and other parameters that alter the trajectories of the two steady states (such as the stimulus intensity, and the settings of W^{EE} and W^{IE}) also affected the generation of Off responses, but these factors were not critical. Moreover, the decreased activity $m_2^E(t)$ during the stimulus is not critical for the generation of the Off response at network level (cf., it is necessary in the post-inhibitory mechanism at cellular level). As can be seen in Figure 4.6d, the amplitude of $v_2^E(t)$ during the stimulus (green trajectory) can be larger compared to no stimulus (blue trajectory).
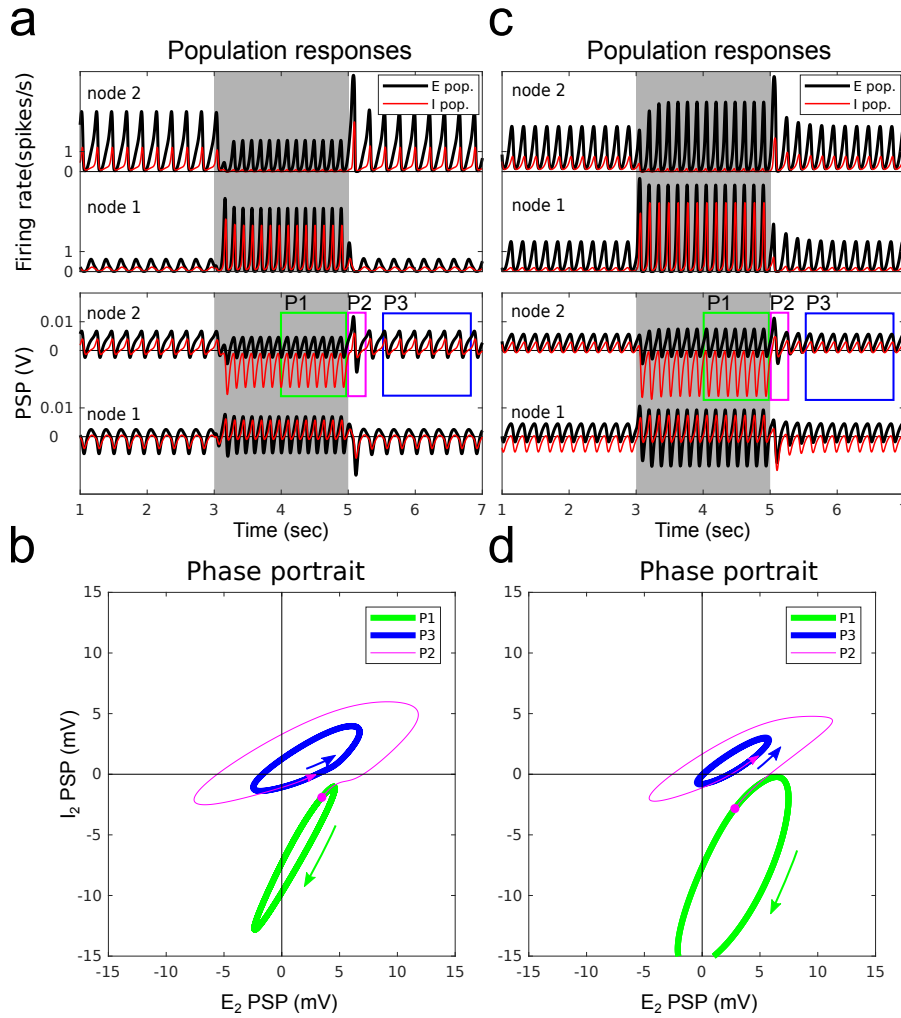
Figure 4.6: Offset detection. (a) In the simulation settings, a prolonged stimulus of 2000 ms is fed to the R node (node 1) in a two-node network. The inter-node connections W are chosen from the W solutions that give rise to Dec-Off responses in the C node (node 2) (see Figure 4.3b). The node responses to the stimulus (the gray period) are shown as firing rates (upper plot) and as PSPs (lower plot). As an Dec-Off response, the time course $m_2^E(t)$ (black curve) shows a lower amplitude during the stimulus and a transient peak at the offset of stimulus. The population I_2 is strongly inhibited during the stimulus, which is reflected by the negative PSP of population I_2 $v_2^I(t)$ (red curve in the green rectangle). The disinhibition is followed by the Off response in population E_2 thereafter (black curve in the magenta rectangle). (b) Phase portraits (P1: during stimulus, P2: offset of stimulus, P3: post stimulus) of node 2. The phase portrait P3 (i.e., when there is only background input) runs counter-clockwise, and the phase portrait P1 (i.e., during the stimulus) shifts downward and runs clockwise, reflecting the strong inhibition of I_2. The phase portrait P2 shows the transient trajectory of transition from P1 to P3. The magenta dot denotes the time of stimulus offset. (c) The simulation settings for a Inc-Off response. The firing rate $m_2^E(t)$ shows higher amplitude during the stimulus and a transient peak at the offset of the stimulus. As in (a), population I_2 is strongly inhibited during the stimulus, which is then followed by the Off response. (d) The phase portrait is similar to (b) except that the amplitude of $v_2^E(t)$ is larger during P1 than P3. The two examples show that the generation of Off responses does not depend on the increased or decreased activity in E_2, but to the inhibition on I_2 during the stimulus.

4.3.3   Robustness of change detection

I have shown that the emergence of change detectors (i.e., the generation of On/Off responses in the two-node network) depends on the inter-node connection patterns, especially the inhibitory connections. I further checked the influence of three factors on the generation of On/Off responses: (1) external input to inhibitory populations, (2) blockage of NMDA receptor channels that alters E/I balance, and (3) synaptic adaptation on connections W^{EE}. More specifically, I assessed how each of these three factors influences the distribution of W solutions in the two-node network in four conditions: (I) default, (II) $W^{IX} = 0$, (III) NMDA-r antagonist, and (IV) synaptic adaptation. The simulation conditions are described below.

1. Since disinhibition played an important role in the generation of both On and Off responses as illustrated in Figure 4.5 and 4.6, I was interested in seeing the contribution of external input to the inhibitory population I_1 . In condition II, the external connection W^{IX} was set to zero in comparison with the default setting $W^{IX} = 0.5W^{EX}$ (condition I).

2. The NMDA-r antagonist MK-801 is found to reduce inhibition during stimulation and thus to reduce the Off responses [Baba et al., 2016]. NMDA-r antagonists are also known to reduce the amplitude of the MMN [Näätänen and Kähkönen, 2009]. In condition III, I mimicked the effect of NMDA-r antagonists by reducing the connection strength of W^{EE} by 25% and reducing W^{IE} by 50%. The difference in reduction applied to the two connections was based on the fact that excitatory synapses on inhibitory neurons are mainly covered by NMDA channels and therefore are more sensitive to NMDA-r antagonists than the excitatory synapses on excitatory neurons [Rujescu et al., 2006]. The setting of external connections remained the same as the default setting. Note that in principle both conditions II and III may be due to NMDA-r antagonists, because they are based on decreased excitatory input to the inhibitory populations. So, if NMDA-r antagonists are indeed the cause of reduced connection strengths to inhibitory populations, the effect in condition II and III should occur simultaneously. Other effects caused by NMDA-r antagonists, such as the changes in NMDA currents, synaptic plasticity and synaptic time constants, were not included.

3. The phenomenon of synaptic adaptation is ubiquitous in the nervous system and has been suggested to be one of the mechanisms underlying deviance

detection. Since the proposed generic deviance detection principle suggests that the deviance-related responses reflect the change detection on regularity representation, it is important to know whether synaptic adaptation promotes the emergence of change detectors. In condition IV, the synaptic adaptation effect was considered. The intra- and inter-node connections W^{EE} were modulated by the synaptic adaptation term a, as described in Equations 3.23 and 3.31. The external input via W^{EX} to the excitatory populations was not affected by synaptic adaptation.

The responses of the two-node network with a range of inter-node connections Ws (as in simulation I in Section 4.2.2) were simulated, and each W was assigned to one of the nine types of responses (Also see Figure 4.3b). Four conditions were tested: (I) the default condition, where synaptic adaptation was not applied, and $W^{IX} = 0.5W^{EX}$, (II) $W^{IX} = 0$, (III) $W^{EE} = 0$ reduced by 25% and $W^{IE} = 0$ reduced by 50%, and (IV) synaptic adaptation applied. To visualize the results, the W solutions of types 1 to 9 were projected to a 2D plane (Figure 4.7a-d). The number of W solutions under the four conditions were summarized in the contingency table (Figure 4.7e-g) and the bar chart (Figure 4.7h).

The bar chart (Figure 4.7h) shows that the number of W solutions of Off types in condition II was reduced compared to condition I. Most of the Off types under condition I became None types under condition II (e.g., Inc-Off $\rightarrow$ Inc-None among 1.25% of the scanned Ws. See the cyan rectangle in Figure 4.7e). This suggests that the external connection W^{IX} is supportive of the generation of Off responses, because the I_1-to-I_2 disinhibition was enhanced due to the external input via W^{IX}.

In condition III the number of W solutions of Off types was reduced, but the number of W solutions of On types was slightly increased compared to condition I (Figure 4.7h). This is in line with experimental results showing that NMDA-r antagonists reduce Off responses but On responses are not affected [Javitt et al., 1996, Umbricht et al., 2000].

In condition IV, the number of W solutions of both On and Off types were greatly increased (Figure 4.7h). Many of the None types under condition I turned into On and Off types under condition III (e.g., 3.28%: Inc-None $\rightarrow$ Inc-On; 1.62%: Dec-None $\rightarrow$ Inc-Off; 2.82%: Dec-None $\rightarrow$ Dec-Off. See the magenta rectangle in Figure 4.7f). This suggests that synaptic adaptation greatly promotes the emergence of change detectors. To see how synaptic adaptation alters the network

responses, Figure 4.8a-c shows three examples of altered responses due to synaptic adaptation. The three examples show typical type transitions from condition I to condition IV.
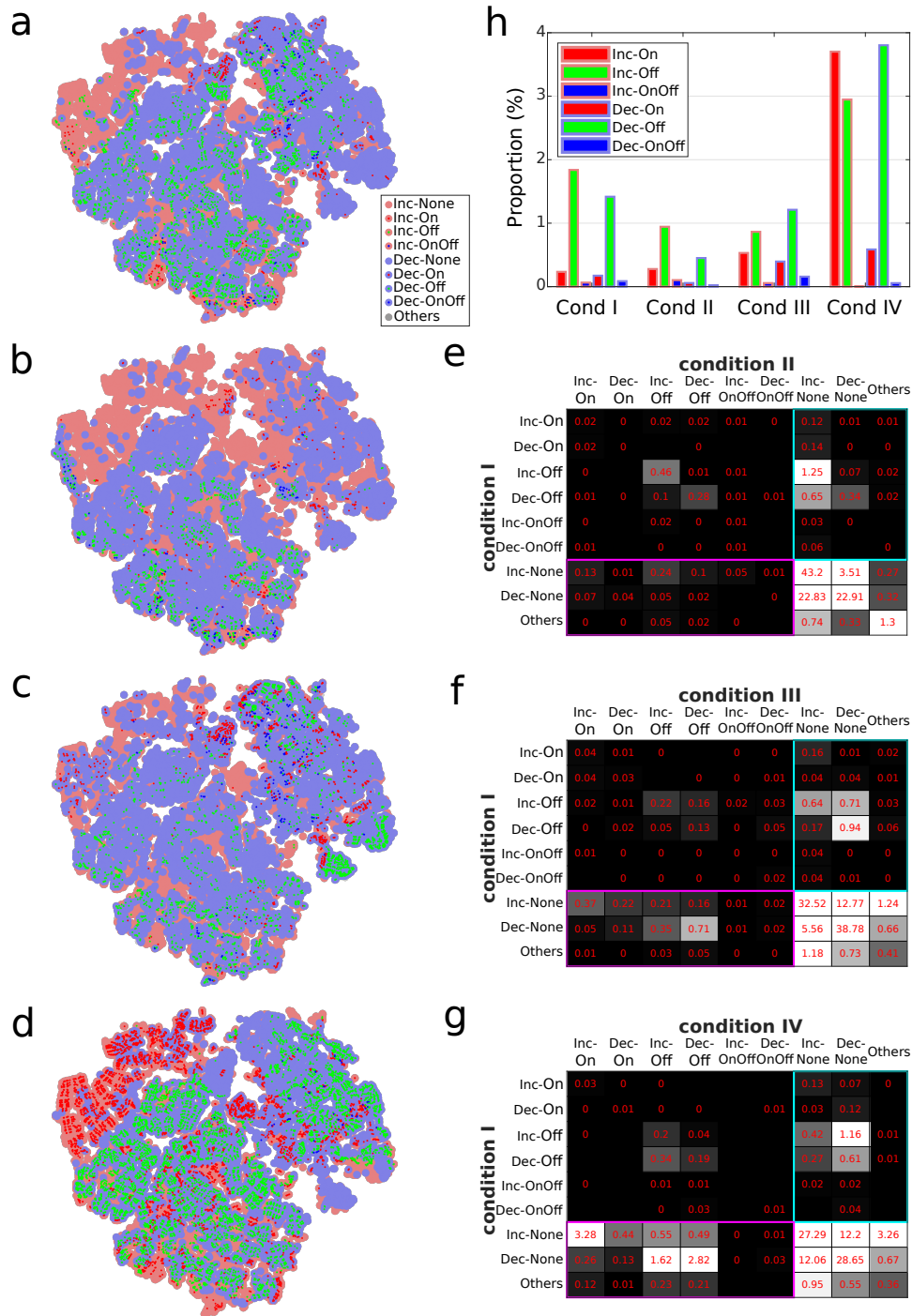
Figure 4.7: The effects of factors influencing the strength of connections W on the occurrence of On/Off responses. The W solutions of On/Off responses projected to a 2D plane under (a) condition I: default, (b) condition II: $W^{IX} = 0$, (c) condition III: NMDA-r antagonist, and (d) condition IV: synaptic adaptation. Dots with different colors and sizes represent different response types. (e) The contingency table of W solutions for condition I vs. II. The value in each cell of the table (in red, with grayscale background) is the number of W solutions over the total number of scanned Ws. A cell without a value means there was no W solution in that case. The cyan and magenta rectangles highlight the W solutions of On/Off types under one condition but not under the other. (f) Condition I vs. III. (g) Condition I vs. IV. (h) The bar chart represents the proportions of W solutions of On/Off types under the four conditions.
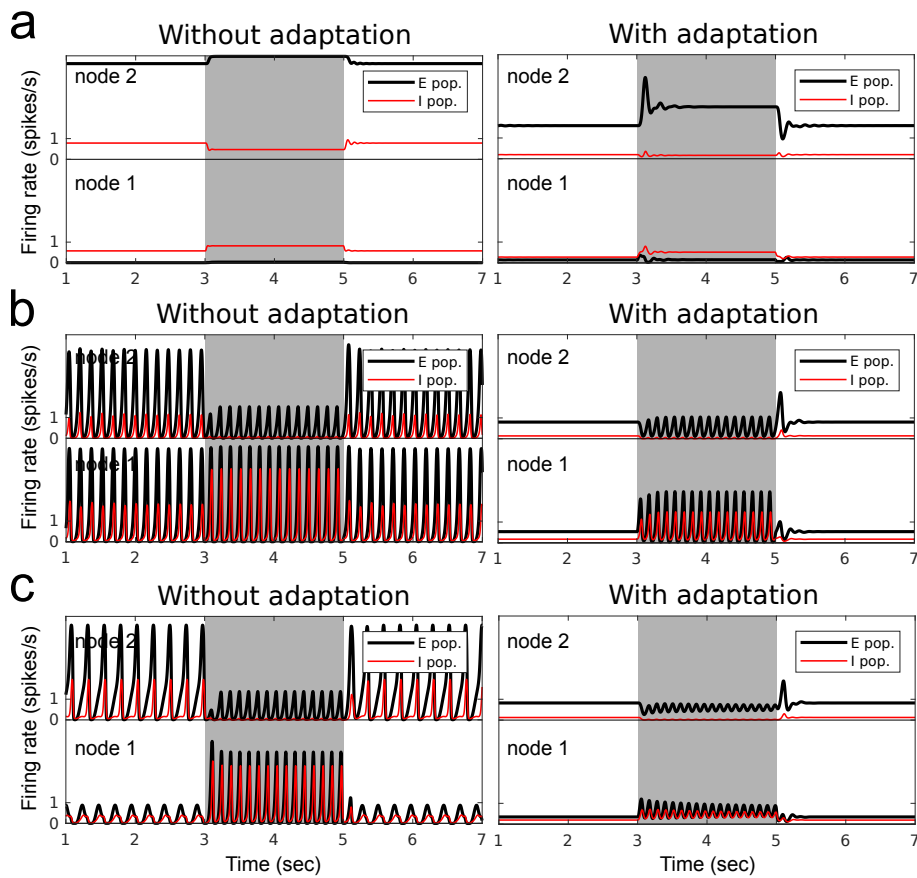
Figure 4.8: Network responses without and with synaptic adaptation. (a) Example of Inc-None type response in population E_2 under condition I (i.e., no adaptation) turning into Inc-On type under condition IV (i.e., when synaptic adaptation is applied to W^{EE}). (b) Example of Dec-None type turning into Inc-Off type. (c) Example of Dec-None type turning into Dec-Off type.

4.3.4   Summary

The recurrent nature of the intracortical wiring makes change detection ubiquitous. Functionally speaking, the ubiquity of change detection across the brain facilitates perceptual representation across the hierarchy. Edge information at all levels, provided by the local change detectors, augments the representational space. Such information compression may also contribute to energy saving. In this sense, the change detectors are more like high-pass filters than comparators that subtract top-down signals from the bottom-up signals. The abundant recurrent wiring patterns in the cortex provide a suitable environment for the emer-

gence of change detectors. I take the diversity of cortical On/Off responses [Chimoto et al., 2002, Deneux et al., 2016, Volkov and Galazjuk, 1991] as an example. Even though these responses could originate from the feedforward mixture of non-cortical On/Off responses at earlier stages such as the thalamus, midbrain, and brainstem, the cortex provides more abundant chances for the emergence of On/Off responses. In simulation I (Section 4.2.2), I demonstrated that various types of On/Off responses can be generated by different inter-node connections (Figure 4.3). In simulation II (Section 4.2.3), I further demonstrated that for a specific connection setting, the difference in input ratios to nodes gives rise to distinct onset and offset FRFs (Figure 4.4). The *W* solutions of On/Off responses projected to the 2D plane (Figure 4.3c) also provided an explanation for the diverse (and spatially clustered) cell responses observed in auditory cortex in awake mice, as shown in Figure 5 in [Deneux et al., 2016]. These results suggest that change detection is a basic and ubiquitous operation in the cortex.

I then studied the generation of On and Off responses. On responses were due to a *transient disinhibition* (i.e., a quick and light inhibition on the inhibitory population of the change detector) before the network reached the steady state (Figure 4.5). Off responses were always associated with *a release from long-lasting disinhibition* (i.e., a long and strong inhibition on the inhibitory population of the change detector) before the network came back to the steady state without the stimulus (Figure 4.6). This is in line with the rebound after inhibition hypothesis [Takahashi et al., 2004, He et al., 1997]. I suggest that the inhibitory to inhibitory connections are a key aspect of change detection.

NMDA-r antagonists dampen the deviance-related responses. I suggest that the NMDA-r antagonists could generally dampen the deviance-related responses through three aspects: (1) voltage-dependency, (2) synaptic plasticity and (3) *E/I* balance (Table 4.2). First, the NMDA-r antagonists block the voltage-dependent NMDA channels and reduce the additional NMDA currents that reflect mismatch signals [Javitt et al., 1996]. Second, the antagonists damage the spike-timing-dependent plasticity (STDP) and hamper the ability of regularity formation [Baba et al., 2016, Javitt and Sweet, 2015, Uhlhaas and Singer, 2010]. Third, the NMDA-r antagonists alter the connection patterns and *E/I* balance. Blocking NMDA receptors leads to decreased activity in the GABAergic interneurons and increased pyramidal excitation, because the GABAergic interneurons are tenfold more sen-

sitive to the NMDA-r antagonists than the pyramidal neurons [Rujescu et al., 2006, Grunze et al., 1996].

The adaptation-based and prediction-based models of MMN (introduced in Sections 2.3.1 and 2.3.2) agree on the voltage-dependency aspect and suggest that the reduced MMN amplitude is due to the reduction in NMDA currents [May and Tiitinen, 2010, Wacongne et al., 2012, Wacongne, 2016]. The prediction-based models also mention the need for STDP to form prediction signals [Wacongne et al., 2012, Wacongne, 2016]. In addition to these two aspects, the simulation results show that the altered E/I balance, as an effect of NMDA-r antagonists, can reduce the emergence of change detectors. In condition III (Figure 4.7c), I reduced the strengths of W^{EE} and W^{IE} by 25% and 50% respectively and recounted the number of each On/Off types in the scanned range of inter-node connection Ws. The number of Off types decreased whereas the number of On types is slightly increased relative to the default setting (Figure 4.7h). These results suggest that the NMDA-r antagonists may in general dampen the deviance-related responses because change responses are reduced.

I cannot draw further quantitative conclusions from the effect of NMDA-r antagonists because the uniform search range of Ws in the simulation is just a simplification. The exact proportion of strength reduction due to NMDA-r antagonists is not available. The settings of 25% and 50% in connection strength reduction in condition III was arbitrary so that a single node still oscillates under a certain range of input intensity, which eliminates the case when the nodes are saturated and no On/Off responses are generated at all. The time constant τ_e, due to the blockage of NMDA channels, was not modified in the simulation in order to focus on the effect of W change.

Table 4.2: A summary of NMDA-r antagonist effect on deviance detection

Stage	Mechanisms		
	Voltage-dependency	STDP	E/I balance
regularity formation		v	v
change detection	v		v

Synaptic adaptation facilitates change detection. Synaptic adaptation is a pervasive short-term plasticity that is considered as a mechanism underlying deviance detection, in the sense that a rare stimulus triggers stronger neural activity via un-

adapted pathways. Given the pervasiveness of synaptic adaptation, I was interested in how it affects the behavior of change detectors in the simulations. In condition IV (Figure 4.7d), the strength of W^{EE} was modulated by short-term adaptation according to the activity of presynaptic excitatory populations. After scanning through the Ws I found that the number of W solutions of both On and Off types were increased compared with the default condition (Figure 4.7g). More specifically, many W solutions of None type turned into On and Off types when synaptic adaptation was applied (as examples in Figure 4.8 show). I suggest that synaptic adaptation facilitates change detection by turning many otherwise None type responses (usually reflected by saturated activity in the excitatory populations) to either On or Off responses.

Chapter 5

Regularity formation

"Once is happenstance. Twice is coincidence. The third time it's enemy action."

– Ian Fleming

5.1 Overview

One can easily recognize repetitive patterns in the acoustic environment such as the sound of flapping wings, cricket, and footsteps on an old wooden floor. The brain automatically finds *temporal patterns* within these acoustic stimuli, where the patterns include *regular timing and order*. Regularity is a type of feature that represents the constancy (or repetitiveness) of its lower level feature. Functionally speaking, regularity provides statistical information of past events, which is useful for better prediction of a future event.

Unlike the feature of change, the feature of regularity is fundamentally more stable. Regularity formation may involve a process of integration, where the neural activities in a certain area are accumulated over time and stay active for a longer period if a threshold is reached, or where the local synaptic patterns adapt to the repetitive feature and become less responsive. In general, the process of regularity formation plays an essential role to reduce the feed-forward propagation of lower-level change responses.

In this chapter, I studied regularity formation, with a focus on the conditions of *regular periodicity* (Section 5.2) and *regular sequence pattern* (Section 5.3). The

former encodes 'when' the stimulus is coming, and the latter encodes 'what' stimulus is coming next. Although the two conditions obviously do not cover the entire space of possible regularities, hopefully their underlying mechanisms can be leveraged and generalized to other types of regularity.

5.2 Regular periodicity

The brain expects the timing of a next event after a few repetitions of periodic events. How the brain processes temporal structure at different timescales relies on different mechanisms. I am particularly in the timescale where the omitted-stimulus response (OSR) is elicited in humans. Therefore, a more specific question was: *How is the periodicity in the range from tens to hundreds of milliseconds encoded and maintained in the primary cortex?* In Section 5.2.1, I reviewed some experimental observations and the existing computational models. In Section 5.2.2, I run simulations and monitor the network dynamics to investigate the important ingredients for encoding regular periodicity. In particular, the effect of the short-term plasticity is examined.

5.2.1 Experimental observations

Experimental evidence showed that the periodicity can be represented spatially. In the midbrain and the auditory cortex, there are maps of periodicity that are orthogonal to the tonotopic arrangement [Langner et al., 2009, Baumann et al., 2011, Barton et al., 2012, Baumann et al., 2015, Brewer and Barton, 2016] (See an fMRI finding in Figure 5.1), suggesting a complete palce coding scheme of tone and periodicity. In addition, periodicity can be maintained in the neural oscillations. As can be seen in the experimental observations, the entrained brain activities seem to persist for a few more cycles at the cessation of the stimulus [Spaak et al., 2014, Forseth et al., 2018] (Figure 5.2). These two properties in the auditory cortex (i.e., *place coding* and *temporal persistence*) contribute to the time perception in the timescale of tens to hundreds of milliseconds.
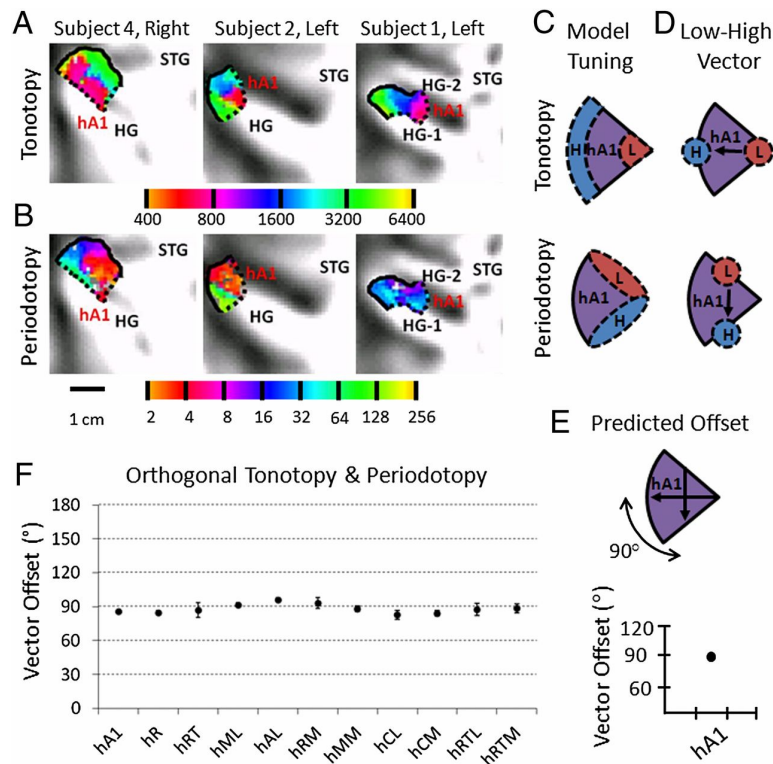
Figure 5.1: Orthogonal tonotopic and periodotopic representation. The fMRI results show the orthogonal gradients in human core and belt auditory cortex. (A) The color-coded tonotopy map ranging from 400 to 6400 Hz. (B) The color-coded periodotopy map ranging from 2 to 256 Hz. (C) and (D) illustrate how the Low-High vectors in each ROI were calculated. (E) The vector offset is calculated as the angle between the two Low-High vectors. (F) All ROIs show vector offset of about 90°. For ROI naming, please refer to the original article. (Figure copied from [Barton et al., 2012])

The computational models that cope with *periodicity representation* usually fall in one of two approaches. *Place coding* is one approach that utilizes the collaboration of different neural populations, each population having fixed properties in terms of time constant or resonance frequency (Figure 5.3a). In this way, the information of periodicity can be represented spatially, in line with the observed periodotopic map shown in Figure 5.1. The problem is that the periodicity representation does not continue when the input stops.

Adaptive tuning is the other type of approach that directly modifies model parameters (e.g., thought synaptic plasticity) so that the model adapts to the input periodicity. In this case, the information of periodicity is encoded in the parameters (e.g.,

synaptic strengths), which results in adaptation, persistence, and even recall of the input periodicity (Figure 5.3b).
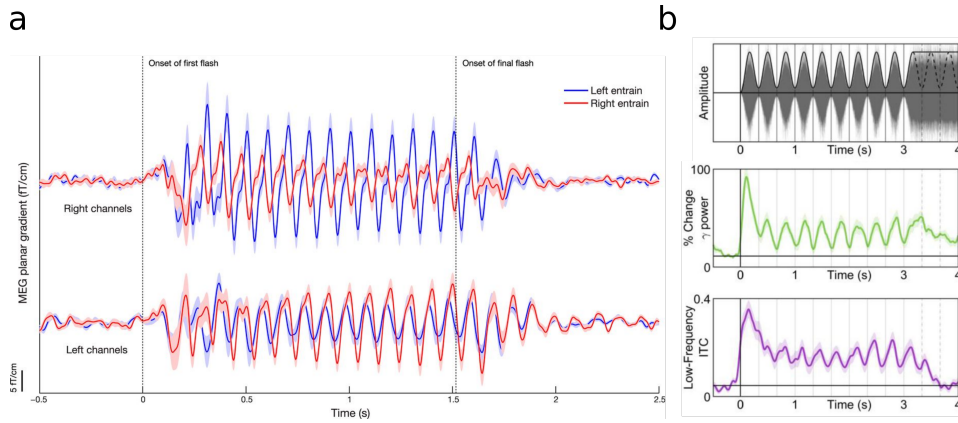

Figure 5.2: Persistent periodicity in the sustained oscillations. (a) In an MEG study, human participants were presented with 10 Hz visual flashes (flash duration, 17 ms; interflash interval, 83 ms; total number of flashes per trial, 16). The periodicity of the entrained ERFs persisted for several cycles after the end of the stimulation (second vertical line). The source analysis suggests the entrained alpha activity was produced locally in the visual cortex. (Figure adapted from [Spaak et al., 2014]) (b) In an intracranial recording study (31 patients, 6580 electrodes, depth probes implanted along the anteroposterior extent of the supratemporal plane), the participants were presented with 3 Hz amplitude-modulated white noise stimulus (upper panel). The entrained phenomena in the 65-115 Hz gamma power (middle panel) and the 2-15 Hz low-frequency inter-trial coherence (lower panel) persist after $t = 3$ s. (Figure adapted from [Forseth et al., 2018])
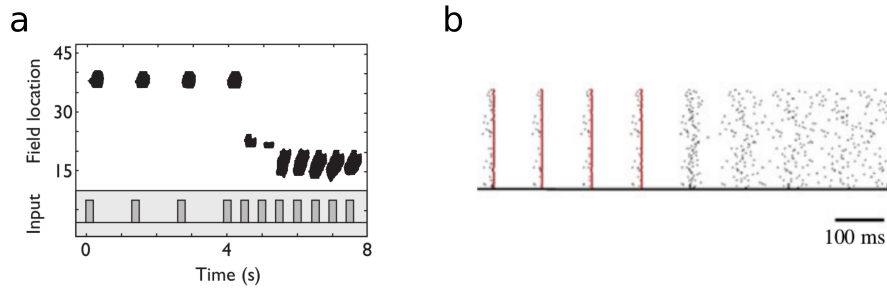
Figure 5.3: Two approaches to encode periodicity. (a) Place coding approach. The selectivity to input SOA is achieved according to the place coding scheme. A field with a discrete set of 50 Wilson-Cowan nodes was simulated. The intra-node synaptic weights (w^{EE}, w^{EI} and w^{IE}) varied across the field so that the resonance frequencies range from 0.8 Hz to 6.67 Hz. (Figure copied from [May and Tiitinen, 2001]) (b) Adaptive tuning approach. A population of interconnected integrate-and-fire neurons ($N = 100$) with spike timing-dependent plasticity (STDP) on the synapses was simulated. This model utilizes a combination of neural diversity and synaptic plasticity to account for precise encoding of input periodicity. Red vertical lines indicate the last 4 pulses in total 12 pulses. (Figure copied from [Thivierge and Cisek, 2011])

5.2.2   Simulation III: encoding periodicity

In the network model, I included both *place coding* and *adaptive tuning* approaches (mentioned in 5.2.1) so that the network properties fulfilled the experimental observations in Figures 5.1 and 5.2.

Simulation settings

Network and input. In the first part of simulation, the network comprised a resonant bank of 16 nodes, where the inter-node connections were initially set to 0 (Figure 5.4a). The intra-node connections w_{kk}^{IE} and w_{kk}^{EI} ($k = 1,2,...,16$) ranged linearly from 0.5 to 4.5, and the background input B_k to each node is selected according to Figure 5.5a so that the nodes worked in the linear range. The external input $x(t)$ contained a prolonged stimulus (duration= 3 s, amplitude= 0.3 spikes/s, rise/fall time= 10 ms) or periodic stimuli (duration= 50 ms; amplitude= 0.3 spikes/s; rise/fall time= 10 ms; SOAs= 100, 140, 180, 220, 260, and 300 ms). The external input was fed to all E populations (i.e., $W^{IX} = 0$). In the second part of simulation, an additional node (change detector) was added to the resonant bank (Figure 5.4b). This node did not receive the external input. The dynamic of the nodes is described

in Equations 3.27, 3.28, 3.29, and 3.30. The settings followed the configurations using Wilson-Cowan-based operators specified in Table 3.2. Detailed settings were specified in Table 5.1
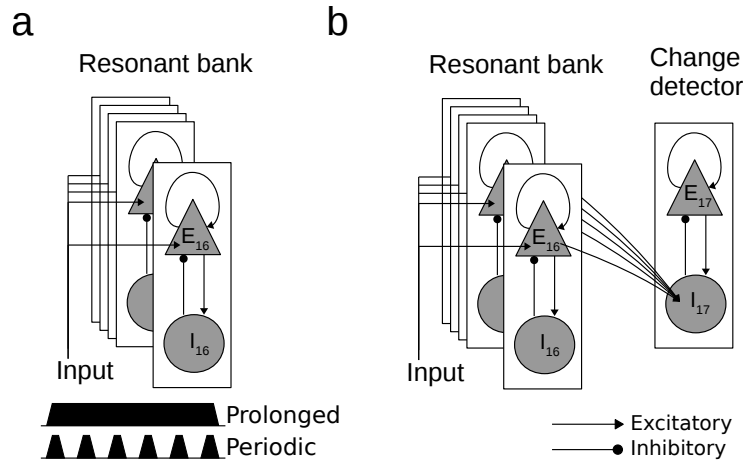

Figure 5.4: Simulation settings. (a) Setting A: the resonant bank comprises 16 nodes. All E populations receive the same external input (a prolonged stimulus or periodic stimuli). There are no inter-node connections initially. The inter-node connection w_{jk}^{EE} may temporarily increase during in-phase oscillation if short-term plasticity is considered. (b) Setting B: a change detector is added to the network. The populations E_{1-16} in the resonant bank provide excitatory inputs to population I_{17}. When the I_{17} activity drops, the originally suppressed population E_{17} shows transient peak with the help of its self-feedback. This thus provides a way to monitor the persistence of the overall E_{1-16} activities by checking the E_{17} peak latency.
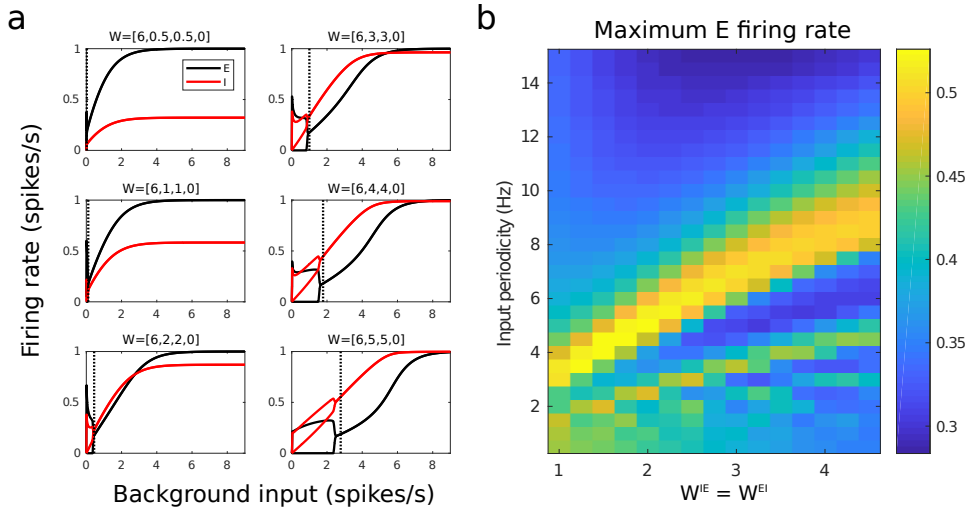
Figure 5.5: Periodicity representation by a resonant bank. (a) The bifurcation diagrams on varying background input. The nodes had different intra-node connection settings ($W = [W^{EE}, W^{IE}, W^{IE}, W^{II}] = [6, x, x, 0]$, where x ranged from 0.5 to 5), which resulted in the different resonance frequencies among the nodes. In the resonant bank, the value of background input to a signal node was decided (vertical dashed lines) so that the steady E and I activities (black and red curves) were not oscillating and were as low as possible. This configuration kept the nodes in the approximately linear working range. The background inputs were then fixed in the simulations. (b) The periodicity representation by the resonant bank under periodic square inputs. The nodes show selectivity to the input periodicity. The periodicity tuning curves of the nodes (the columns in the plane) were color coded. In this plot, the inter-node connections in the resonant bank were set to 0.

Short-term plasticity. Synaptic adaptation on W^{EE} (as described in Equation 3.31) was considered throughout this simulation. In addition, another plasticity term p^{EE} was applied also on W^{EE} in the resonant bank. The general form of plasticity term p_{kj} (j≠k) described in Equation 3.32 is rephrased into below Equation 5.1, where p_{kj}^{EE} strengthen the efficacy of w_{kj}^{EE}. This plasticity rule adjusted the binding between the nodes in the resonant bank so that the group activity maintains a stable representation of input periodicity. The plasticity term p_{kj}^{EE} increases if the covariance $Cov_{j,k,\Delta t}(t)$ between $m_j^E(t)$ and $m_k^E(t)$ from time $t - \Delta t$ to t is positive, and otherwise decreases gradually back to zero. The weight mask g_{kj}^{EE} considered the effectiveness of plasticity as a function of the distance between nodes j and k. For 16 nodes in the resonant bank, the weight mask followed the Gaussian function $exp\left(-(j-k)^2/2\sigma^2\right)$, where σ was set to 4. Since the resonance frequency increases monotonically with the node index in the resonant bank, the weight masks avoid the binding between two nodes with distinct resonance frequencies. This whole setting of a resonant bank enables the resonance to be sus-

tained after the due-time so that the change detector would rise when the sustained resonance drops.

$$\dot{p}_{kj}^{EE}(t) = \frac{-p_{kj}^{EE}(t)}{\tau_p^{EE}} + \kappa_p^{EE} \cdot g_{kj}^{EE} \cdot \left[Cov_{j,k,\Delta t}(t)\right]_+ \tag{5.1}$$

Simulated MEG. The simulated MEG signals were generated according to Section 3.6. The node weights r_k ($k = 1$ to 16) were all set to 1. The contribution of positive and negative ion currents ($r_{current1}$ and $r_{current2}$) were all set to 0.5.

Simulation results

Short-term plasticity (STP) contributes to the persistence in the simulated MEG signals. In the simulation setting A, the nodes stay unconnected (initial off-diagonal $W^{EE} = 0$) when the STP on W^{EE} in the resonant bank ($N = 16$ nodes) was not considered. Even though some nodes resonate with the periodic stimuli and keep oscillating for a while after the end of stimulus, the simulated MEG signals do not persist (Figure 5.6). When the STP was considered, the W^{EE} among the resonating nodes was strengthened. These connected nodes create stronger momentum in the corresponding resonance frequency and therefore the simulated MEG signals persist (Figure 5.7).

Table 5.1: Configurations for simulation III: encoding periodicity

Par.	Value (Node 1-16)	Value (Node 17)	Unit
Impulse response function			
τ_e	20	20	ms
τ_i	60	60	ms
Intra-node connections			
w_{kk}^{EE}	6	6	1
w_{kk}^{IE}	0.5 to 4.5	4.5	1
w_{kk}^{EI}	0.5 to 4.5	4.5	1
w_{kk}^{II}	0	0	1
Background inputs			
B_k	$w_{kk}^{IE} \cdot w_{kk}^{EI}/9$	2	spikes/s
External connections			
w^{EX}	1	0	1
w^{IX}	0	0	1

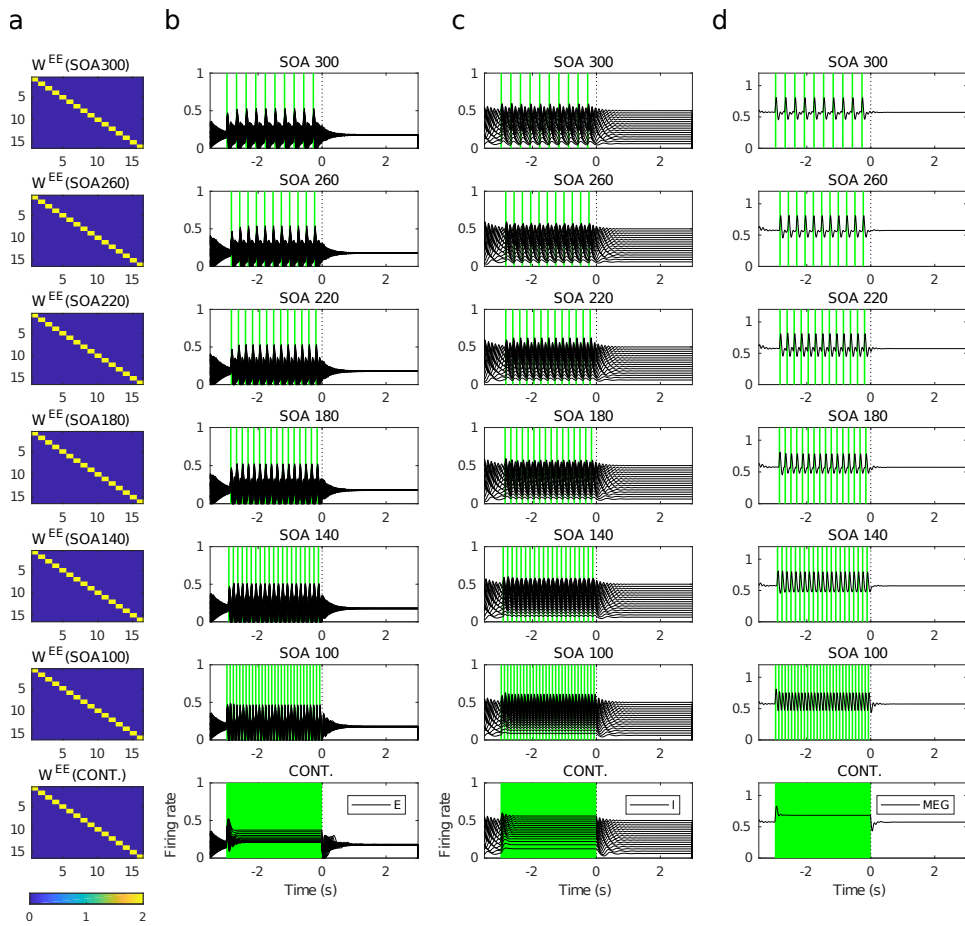| Par.	Value	Unit
Synaptic adaptation		
τ_a	600	ms
κ_a	20	1
Inter-node connections ($j = 1, 2, ..., 16; k = 1, 2, ..., 16$)		
$[w_{kj}^{EE}, w_{kj}^{IE}, w_{kj}^{EI}, w_{kj}^{II}]$	[0, 0, 0, 0]	1
Inter-node connections ($j = 1, 2, ..., 16; k = 17$)		
$[w_{kj}^{EE}, w_{kj}^{IE}, w_{kj}^{EI}, w_{kj}^{II}]$	[0, 3, 0, 0]/16	1
$[w_{jk}^{EE}, w_{jk}^{IE}, w_{jk}^{EI}, w_{jk}^{II}]$	[0, 0, 0, 0]	1

Figure 5.6: Network behavior (The 16-node resonant bank, without short-term plasticity on W^{EE}) under prolonged stimulus (CONT.) and periodic stimuli (SOA= 100,140, 180, 220, 260 and 300 ms; tone duration= 50 ms). (a) The inter-node W^{EE} is 0 through out the simulation time. (b) The firing rate of populations E_{1-16} in response to the stimuli (green). (c) The firing rate of populations I_{1-16}. (d) The simulated MEG signals.
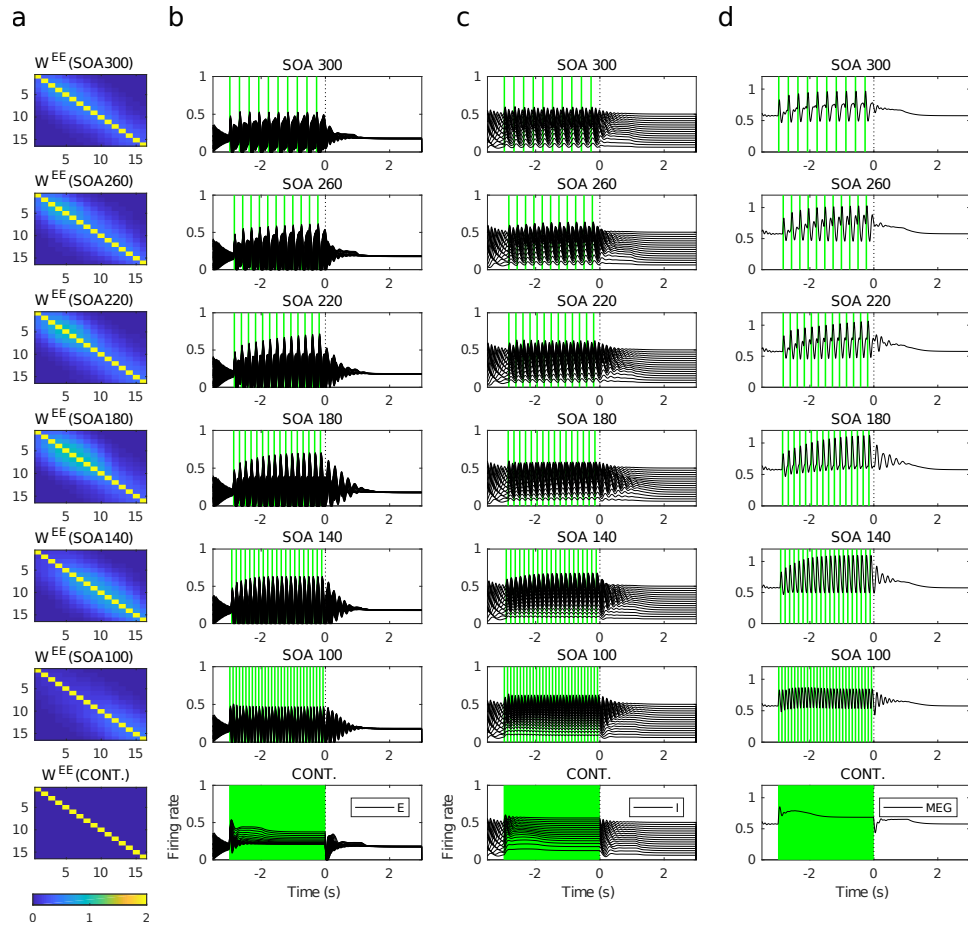
Figure 5.7: Network behavior (The 16-node resonant bank, with short-term plasticity on W^{EE}) under prolonged stimulus (CONT.) and periodic stimuli (SOA= 100,140, 180, 220, 260 and 300 ms; tone duration= 50 ms). (a) The inter-node W^{EE} at $t = 0.5$ s. (b) The firing rate of populations E_{1-16} in response to the stimuli (green). (c) The firing rate of populations I_{1-16}. (d) The simulated MEG signals.

STP contributes to the linear relationship between the sustained activities and the input SOAs. According to generic deviance detection principle, the fall of the sustained activities is followed by the rise of OSR (See illustration in Figure 2.1), so the fall time of the sustained activities affects the latency of OSR. Since the latency of OSR is linear to input SOA, I have investigated the relationship between the sustained activities of the resonant bank and the input SOAs. In simulation setting B, I connected a change detector to the resonant bank and examine the peak latencies. The result shows that the peak latencies of the change detector are linear to the input SOAs only when the STP is considered (Figure 5.8). I found

that the interval between I_{17} trough and E_{17} peak (the red and blue triangles in Figure 5.8cd) is relatively constant, which means that the I_{17} trough determines the peak latency. Since the population I_{17} only receives the excitatory inputs from populations E_{1-16} because population E_{17} is inactive, the I_{17} activity reflects the collective contribution of the resonant bank (i.e., populations E_{1-16}). I examined how the STP reshapes E_{1-16} activities which then affects I_{17} activities. In 'STPon' condition, the I_{17} firing rate increases linearly with the input SOAs, which is not seen in 'STPoff' condition (Figure 5.9). Therefore, I suggest that the STP increases the collective force of the resonant bank, which results in the linear relationship between I_{17} trough latencies and the input SOAs. In sum, the input SOA can be implicitly encoded in the connection matrix through short-term plasticity, and be explicitly represented in E_{1-16} activities so that the change in input periodicity can be detected.
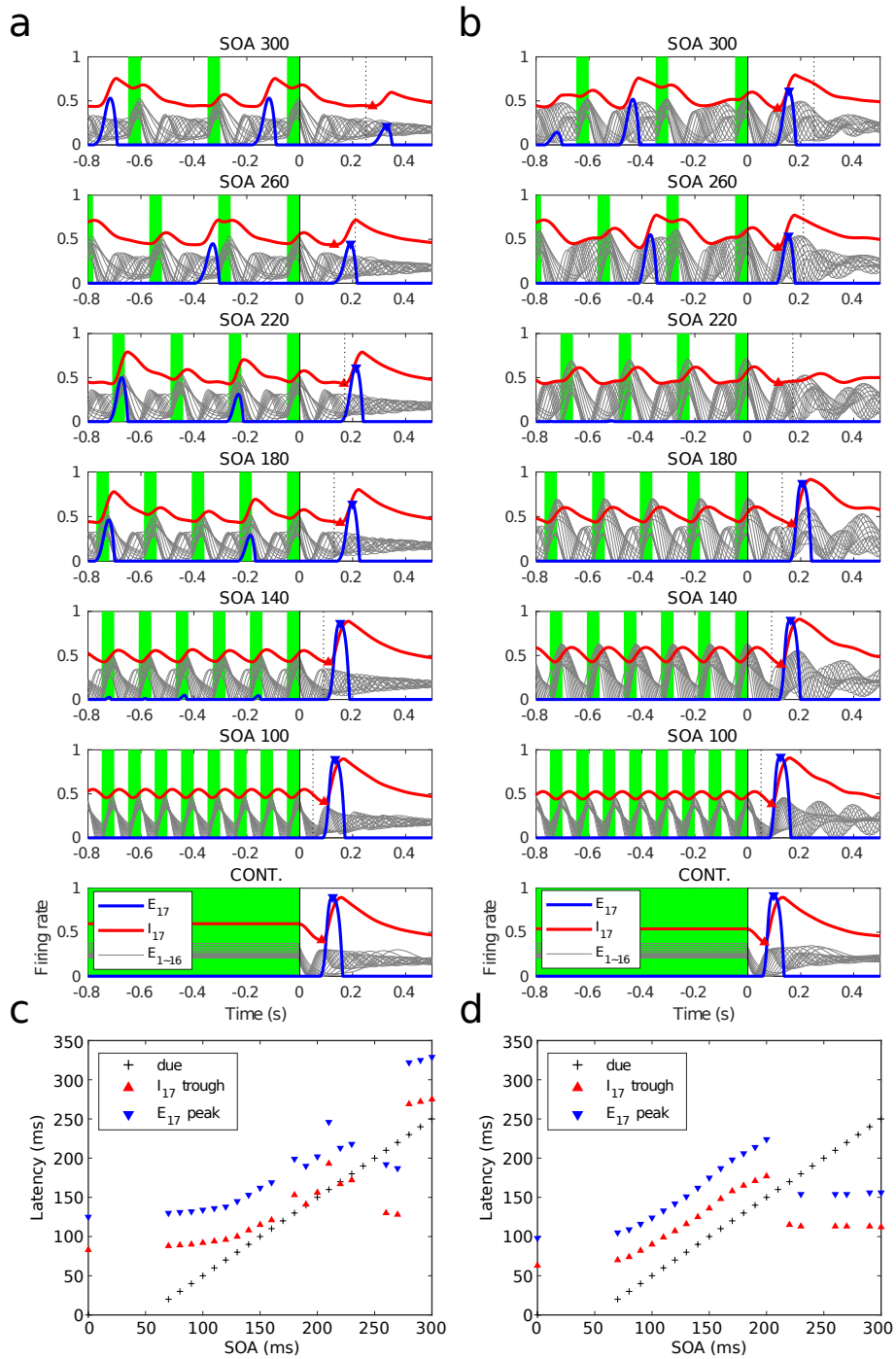
Figure 5.8: The response of the change detector (Node 17) with/without short-term plasticity. (a) In the 'STPoff' condition, the STP on W^{EE} in the resonant bank was not considered. Times are aligned at the end of stimulus (black vertical lines), and the due time (vertical dashed lines) indicates the onset of omission. The I_{17} troughs and the E_{17} peaks are marked by the red and blue triangles, respectively. (b) In the 'STPon' condition, the STP is considered. (c) and (d) The latencies of I_{17} trough and E_{17} peak in 'STPoff' and 'STPon' conditions, respectively. (Setting: $\tau_e = 20$ ms, $\tau_i = 100$ ms)
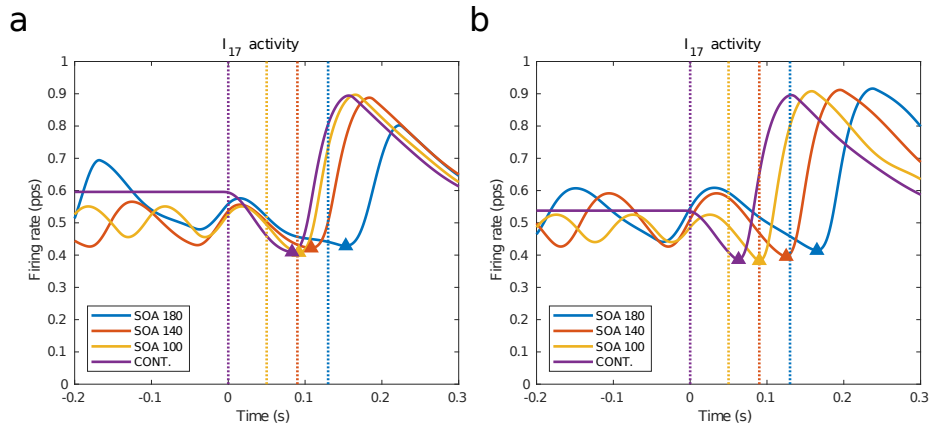
a

b

Figure 5.9: The activities of I_{17} in the change detector under (a) condition 'STPoff' and (b) condition 'STPon'. The activities in response to CONT., SOA 100, 140 and 180 are overlaid. The colored vertical lines indicate the due time. The color triangles indicate the I_{17} troughs. (Setting: $\tau_e = 20$ ms, $\tau_i = 100$ ms)

The time constant τ_i controls the capacity of memory trace. I vary the time constant τ_i and see how it affects the latency-SOA plot (Figure 5.10). The result shows that larger τ_i contributes to a longer capacity of the memory trace. The capacity cannot be improved by simply increasing the learning rate of STP. One thing to note, the learning rate of STP (when above a certain value) does not lead to a steeper slope in the latency-SOA plot. However, the learning rate of STP should not be set too large, otherwise the resonant bank saturates and input periodicity cannot be represented.
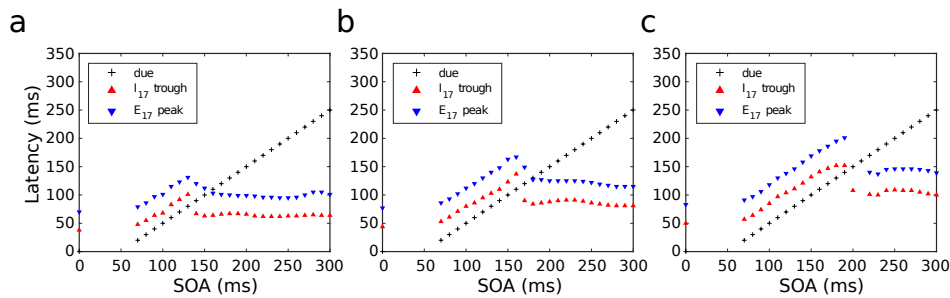
a

b

c

Figure 5.10: Effect of the time constant τ_i. The latencies of I_{17} trough and E_{17} peak are plotted under the conditions (a) $\tau_i = 40$ ms, (b) $\tau_i = 60$ ms and (c) $\tau_i = 80$ ms. The time constant $\tau_e = 20$ ms is kept the same. The time constants τ_e and τ_i are applied to all nodes in the network.

5.3   Regular sequence pattern

The brain expects not only 'when' but also 'what' event comes next after a few repetitions of a sequenced pattern. In general, the expectation forms gradually through repetitions and collapses on a deviant event. Here the question regarding a regular sequence pattern was: *How does the primary cortex encode and represent the regularity of the repeated pattern?* In Section 5.3.1 I review some experimental observations and explanations. In Section 5.3.2 I run simulations and monitored the network dynamics (including the post-synaptic potential, firing rate and connection strength) that mimicked the experimental observations. I summarized important ingredients for encoding regular sequence pattern.

5.3.1   Experimental observations

Neural correlates of regularity formation in complex temporal patterns (found across sensory modalities and species) exhibit several properties. In a two-photon calcium imaging study [Homann et al., 2017] where regular image sequences were repeatedly displayed to the mice, the results showed that the excitatory neurons in layer 2/3 of primary visual cortex soon get adapted and reached a steady state within a fixed decay time constant (ca. 2.1 cycles) (Figure 5.11). In an MEG study [Barascud et al., 2016] where the regular (REG) and random (RAND) sequences of tone pips are played to human participants, the results showed several transient and slow responses during the stimuli (Figures 5.12 and 5.13). The transient responses include the On and Off responses at the stimulus onsets and offsets, and a mismatch response at the switch in REG-RAND condition (but not in RAND-REG condition), which are related to change detection. The slow responses include the shifts toward different steady levels, which are more related to the process of regularity formation. To sum up, four observed properties provide clues and constraints for modeling the process of regularity formation:

1. **Some excitatory neurons get adapted.** This can be seen in the excitatory neurons in layer 2/3 of primary visual cortex (Figure 5.11).

2. **The regularity is formed within a few cycles.** This can be seen in the fixed decay time constant (about 2.1 cycles) in mice (Figure 5.11), and the gradually increased root mean square (RMS) of MEG responses in REG conditions (about 1.5 cycles) in humans (Figure 5.12A).

3. **Neural activities reflect degree of regularity (predictability).** As shown
   in Figure 5.12B, the group RMS of MEG responses are higher in REG than
   in RAND conditions (i.e., group RMS: $REG_n > RAND_n$, where n is alpha-
   bet size[1]). As shown in Figure 5.12A and C, the group RMS show depen-
   dence on the alphabet size n (i.e., group RMS: $REG_n > REG_{n+1}$; $RAND_n >$
   $RAND_{n+1}$).

4. **The regularity collapses quickly once a violation happens.** This can be
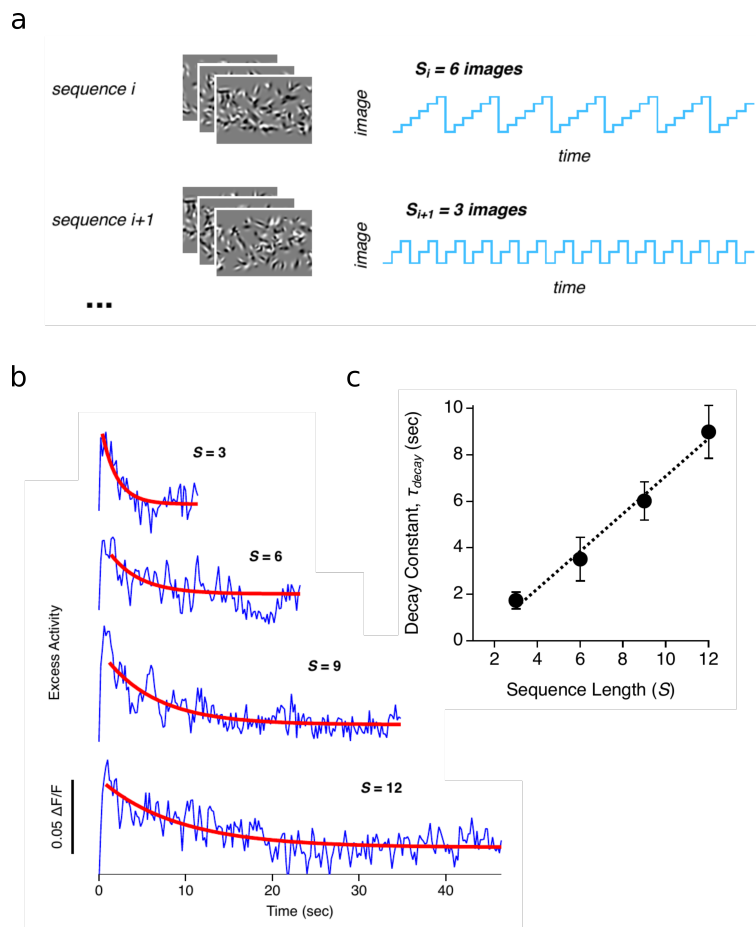   seen in the REG-RAND conditions in Figure 5.13a and b.

Figure 5.11: Two-photon calcium imaging observations under regular sequence pattern.
(a) The visual stimuli used in the experiment. Sequence length is varied in 3, 6, 9 and 12.
Each sequence is repeated 20 times. All images have the same duration of 300 ms. (b)
Population average activity (blue) and the fitted exponential curve (red) for each sequence
length. (c) The decay time ($\tau_{decay} = 2.1 \pm 0.3$ cycles) scales linearly with the total temporal
duration of the sequence. (figure copied from [Homann et al., 2017])

---

[1]Alphabet size is the number of different tones in the random or regular tone sequences.
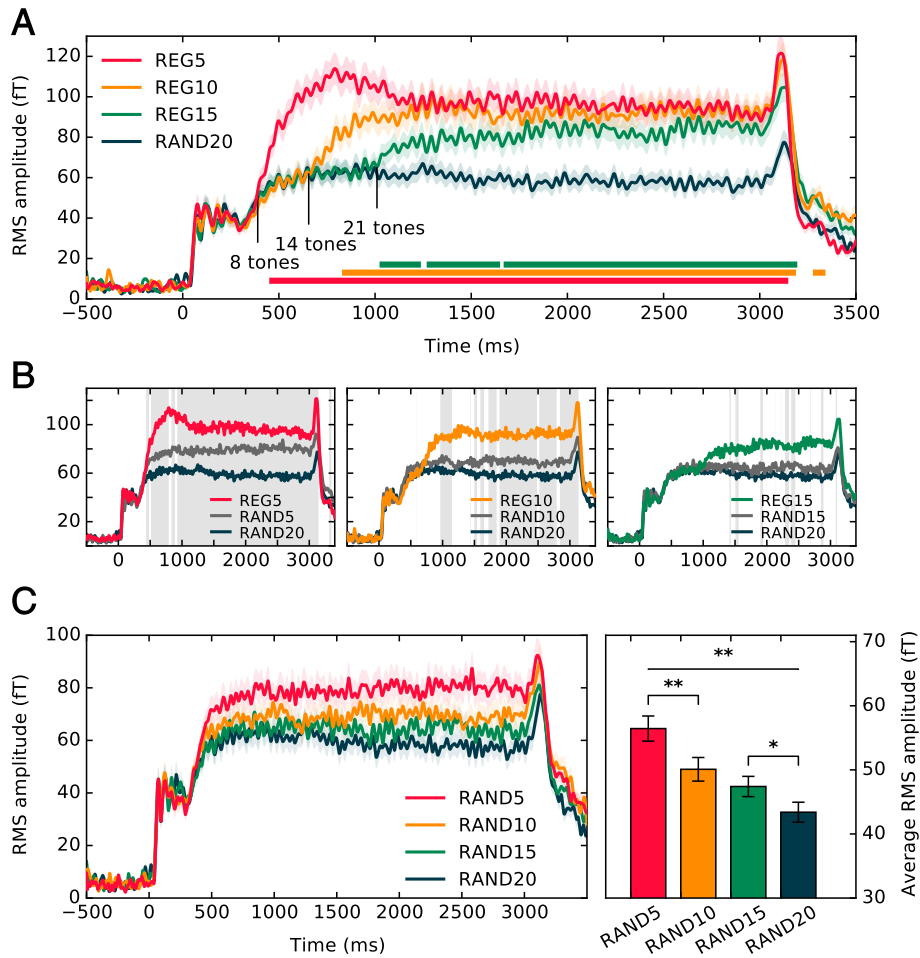
Figure 5.12: MEG responses to regular (REG) and random (RAND) sequences. (A) Group RMS of MEG responses to the regular sequences with alphabet size $n = 5$, 10, and 15 tones (REG5, REG10, REG15) and random sequence with alphabet size $n = 20$ tones (RAND20). The sequences started at $t = 0$ and ended at $t = 3000$ ms. The horizontal color lines indicate the significant differences between the REG and RAND conditions. (B) Group RMS of MEG responses in REG5/10/15 and RAND5/10/15/20 conditions. The gray areas indicate the significant differences between REG5/10/15 and RAND5/10/15, respectively. (C) Group RMS in RAND5/10/15/20 (left) and average RMS amplitude over $t = 0$ to 3000 ms (right). (Figure copied from [Barascud et al., 2016])
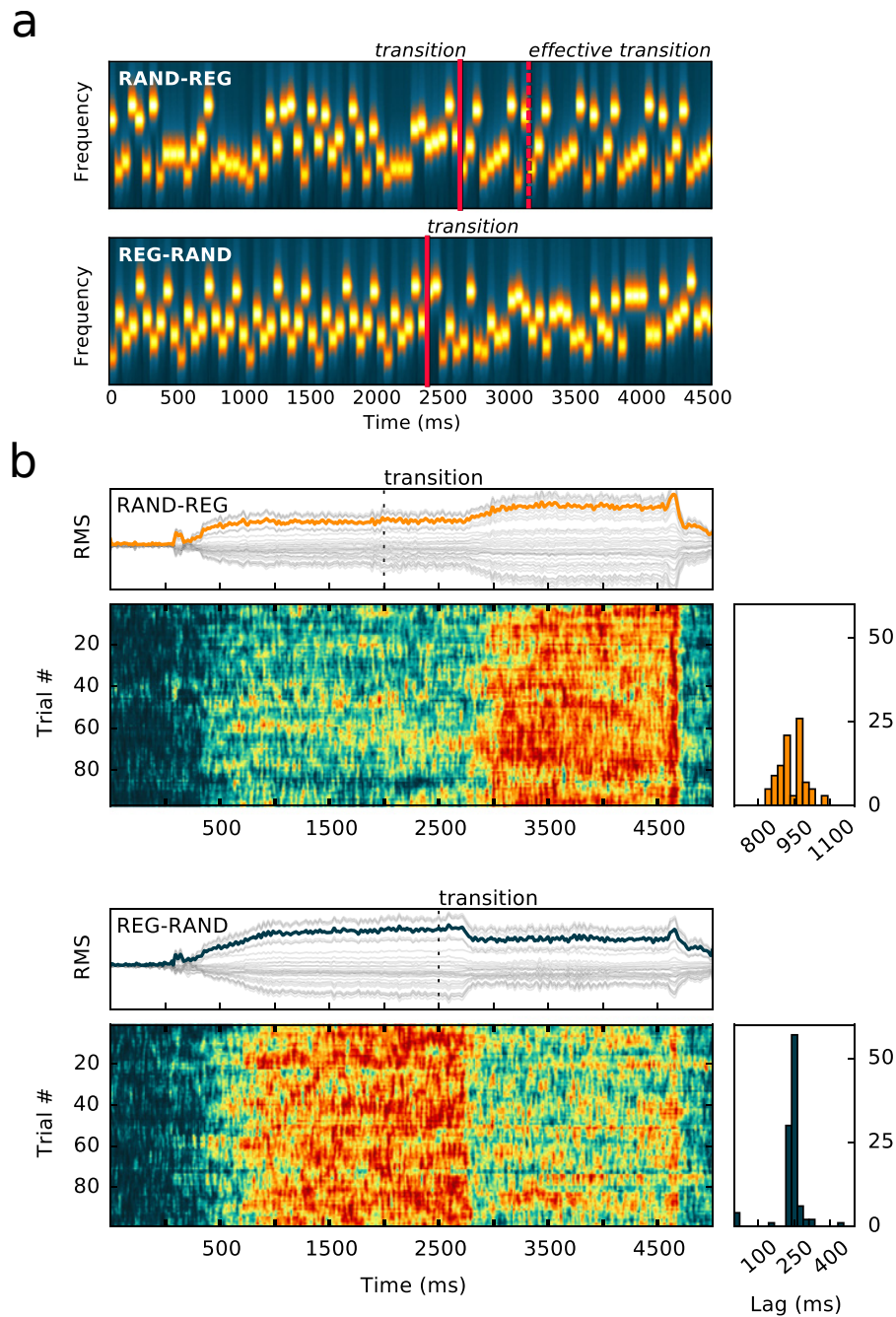
Figure 5.13: MEG responses in RAND-REG and RAND-REG conditions. (a) The stimulus input contained sequences of tone pips (duration= 50 ms) out of 10 tones in random order (RAND) and regular order (REG). The red lines indicate the time of transition. The red dashed line indicates the onset time of a second regular sequence. (b) Single trial data from a representative subject (upper panel: RAND-REG; lower panel: REG-RAND). A single-trial RMS time course (yellow or blue curve) was computed from 40 channels (gray curves). For visualization, each single-trial RMS has been temporally smoothed (moving average window 16.6 ms) and normalized between 0 and 1. The right plots show the histograms of estimated lags of transition responses, where an estimated lag was the lag that gave the maximum cross-correlation value between a single-trial RMS with a Heaviside step function. (Figures copied from [Barascud et al., 2016])

5.3.2 Simulation IV: encoding regular patterns

In this simulation, I investigated how a network encodes and represents the regularity of sequence pattern by reproducing the MEG results in [Barascud et al., 2016]. Three short-term plasticity rules were tested, and the network dynamics were monitored. The four experimentally observed properties regarding regularity formation mentioned in Section 5.3.1 were addressed in the simulation results.

Simulation settings

Network and input. The network comprised 60 nodes (i.e., 15 resonant banks of 4 nodes, Figure 5.14a). A resonant bank consisted of 4 nodes with the intra-node connections $w_{kk}^{IE} = w_{kk}^{EI} = \{0.2, 2.3, 4.4, 6.5\}$, resulting in different resonance frequencies ω_i ($i = 1,2,3,4$). Each resonant bank is selective to one of the input tones $x_q(t)$ ($q = 1,2,...,15$). The input was fed to the E populations (i.e., $W^{IX} = 0$). The nodes with the same resonance frequency were are all-to-all connected ($w_{kj}^{EE} = 0.05$ and $w_{kj}^{IE} = 1.5$, where $j \neq k$). The nodes within the resonant bank were not connected for simplicity[2] The simulation inputs (Figure 5.15) contained regular and random sequences (i.e., REGn-RANDn and RANDn-REGn, where the alphabet size $n = 2,3,...,15$). The dynamic of the nodes is described in Equations 3.27, 3.28, 3.29, and 3.30. The settings followed the configurations using Wilson-Cowan-based operators specified in Table 3.2. Detailed settings were specified in Table 5.2

---

[2]This ignored the interaction between nodes with different resonance frequencies. The purpose of including the heterogeneity in resonance frequency was only to provides a tolerance to the various input SOAs and to render more natural simulated MEG signals.
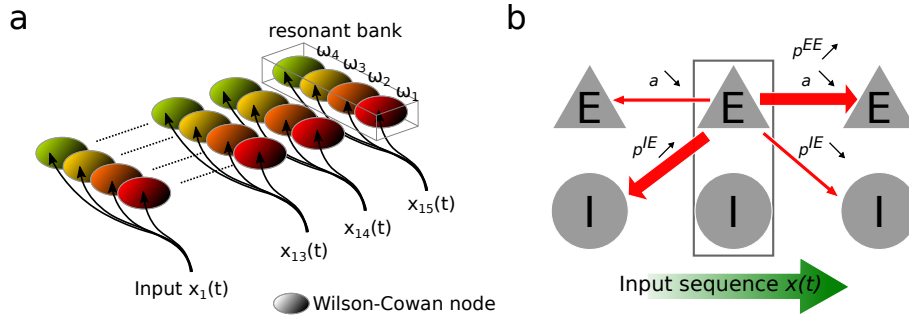
Figure 5.14: Settings of the network and short-term plasticity. (a) The network comprised 60 Wilson-Cowan nodes. The inputs $x_1, x_2, ..., x_{15}$ representing the 15 tones were fed to the resonant banks. A resonant banks comprised 4 nodes with different resonance frequencies $\omega_1 < \omega_2 < \omega_3 < \omega_4$. The resonance frequency was controlled by intra-node connections w_{kk}^{IE} and w_{kk}^{EI} (similar as in Figure 5.5). The inter-node connections are not shown here. (b) Three short-term plasticity terms a, p^{EE}, and p^{IE} that were applied to the connections w^{EE} and p^{IE}. See rules in the context.
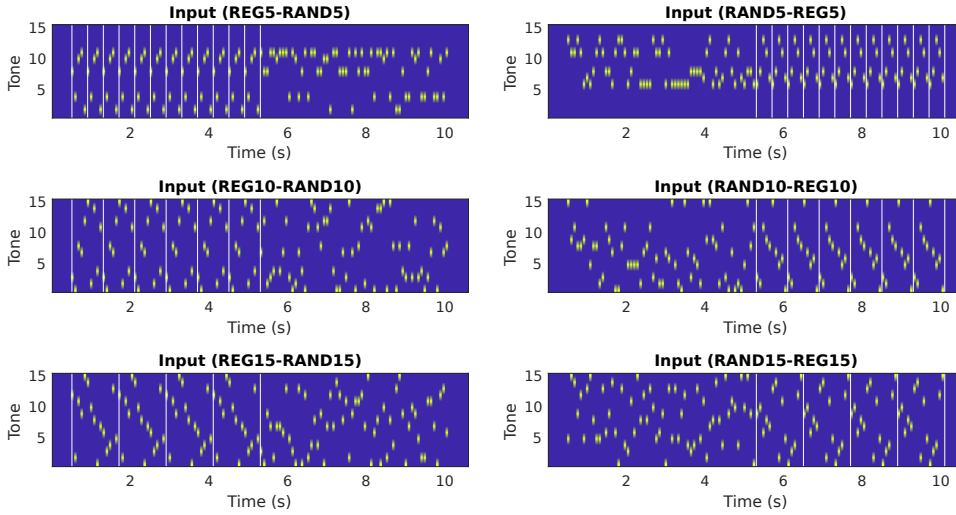
Figure 5.15: Simulation inputs. The input is generated according to the protocol in [Barascud et al., 2016]. In the simulation, the alphabet size is set from 2 to 15 (only 5, 10 and 15 shown here). Tone duration is 80 ms, with no gap between tones. The inputs start with either regular (REG) or random (RAND) patterns. The white vertical lines indicate the starts and ends of repetitive patterns.

Short-term plasticity. As illustrated in Figure 5.14b, the strength of W^{EE} was modulated by two short-term plasticity terms as $a_{kj} \cdot p_{kj}^{EE} \cdot w_{kj}^{EE}$, where the depression term $a_{kj} \in [0, 1]$ decreases from 1 according to the pre-synaptic activity $m_j^E(t)$ (Equation 3.31), and the facilitation term $p_{kj}^{EE} \in [1, x]$ (x: arbitrary upper limit)

increases from 1 when the external inputs $x(t)$ activate the nodes in sequence order $j \rightarrow k$ (Equation 5.2). The strength of W^{IE} was modulated by one short-term plasticity term as $p_{kj}^{IE} \cdot w_{kj}^{IE}$, where the depression/facilitation term $p_{kj}^{IE} \in [0, y]$ (y: arbitrary upper limit) decreases from 1 in sequence order $j \rightarrow k$, and increases from 1 in sequence order $k \rightarrow j$ (Equation 5.3).

$$\dot{p}_{kj}^{EE} = \frac{1 - p_{kj}^{EE}}{\tau_p^{EE}} + \kappa_p^{EE} p_{jk}^{EE} \left[x_j(t - \Delta t) x_k(t) \right] \tag{5.2}$$

$$\dot{p}_{kj}^{IE} = \frac{1 - p_{kj}^{IE}}{\tau_p^{IE}} + \kappa_p^{IE} \left(1 + |1 - p_{kj}^{IE}| \right) \left[x_k(t - \Delta t) x_j(t) - x_j(t - \Delta t) x_k(t) \right] \tag{5.3}$$

So far, the rules for the short-term plasticity terms p_{kj}^{EE} and p_{kj}^{IE} are rather function-driven because they consider both pre- and post-synaptic activities[3]. Basically, the short-term plasticity is use-dependent synaptic plasticity that regulates neurotransmitter release, and it usually only dependents on the pre-synaptic activity. However, there is evidence showing that the post-synaptic dendrite also releases retrograde messages that influence the release of pre-synaptic neurotransmitter [Zucker and Regehr, 2002, Regehr, 2012, Catterall et al., 2013]. I took such dependency on post-synaptic levels of $[Ca^{2+}]_i$ into account and assume that certain forms of short-term plasticity that depends on both pre- and post-synaptic activities exist. This assumption seems to be necessary to achieve the selectivity of sequence order within a few repetitions. More sophisticated rules are still needed.

Simulated MEG. The MEG signal is simulated as described in Equation 3.33 but with an additional short-term plasticity term p_{kj}^{EE} that was introduced to modulate w_{kj}^{EE} (Equations 5.4 and 5.5).

$$R_k(t) = \sum_{j=1}^{N} \left[r_{current1} \cdot a_{kj} p_{kj}^{EE} w_{kj}^{EE} m_j^E(t) + r_{current2} \cdot w_{kj}^{EI} m_j^I(t) \right] \tag{5.4}$$

$$R(t) = \sum_{k=1}^{N} r_k \cdot R_k(t) \tag{5.5}$$

---

[3]The terms p_{kj}^{EE} and p_{kj}^{IE} depend on the sequence order of input $x(t)$ rather than the pre- and post-synaptic activities $m(t)$. This is for stability concern, otherwise the values of p_{kj}^{EE} and p_{kj}^{IE} would be very jumpy and unpredictable.

Table 5.2: Configurations for simulation IV: encoding regular patterns

Par.	Value	Unit
Impulse response function		
τ_e	20	ms
τ_i	100	ms
Intra-node connections		
w_{kk}^{EE}	6	1
w_{kk}^{IE}	0.5 to 4.5	1
w_{kk}^{EI}	0.5 to 4.5	1
w_{kk}^{II}	0	1
Background inputs		
B_k	0	spikes/s
External connections		
w^{EX}	1	1
w^{IX}	0	1
Inter-node connections (Nodes j,k having the same resonance frequency		
$[w_{kj}^{EE}, w_{kj}^{IE}, w_{kj}^{EI}, w_{kj}^{II}]$	$[0.05, 1.5, 0, 0]$	1
Short-term plasticity		
τ_a	600	ms
κ_a	20	1

Asymmetry index (AI). The two short-term plasticity terms p_{kj}^{EE} and p_{kj}^{IE} cause asymmetric matrices P^{EE} and P^{IE}. I used AI^{EE} and AI^{IE} to quantify the change in connection strength induced by the two plasticity rules. The *AI* was calculated as the ratio of the 1-norm of the asymmetric part to the 1-norm of the symmetric part (Equation 5.6), where the diagonal elements in *P* were set to 0.

$$AI = \frac{||P - P^T||}{||P + P^T||} \tag{5.6}$$

Simulation results

Some excitatory neurons get adapted to the repetitive stimuli. This phenomenon can be reproduced in the simulation by including the adaptation term a that was applied to the connection W^{EE} (not shown in the simulation results). The value of a_{kj} ($k = 1, 2, ..., N$) depends on the presynaptic firing rate $m_j^E(t)$, which especially affects the efficacy of self-feedback w_{jj}^{EE} of Node j. The population E_j responds less if it was previously activated by the input $x(t)$ because a_{jj} has not fully recovered yet. Therefore, the decay time of E activities depends on the *length of regular sequence*, which explains the observation found in Figure 5.11.

Since I also included another facilitation term p_{kj}^{EE} on the connection w_{kj}^{EE}, the pure adaptation phenomenon became less clear in the simulation results. The most prominent phenomenon contributed by the adaptation term a_{kj} was the On responses, because the adaptation effect quickly attenuates the initially synchronous firing in E activity by the stimulus. The On responses can be seen in the E activities (upper plots in Figure 5.16b) and the simulated MEG signals (Figure 5.17a) at the stimulus onset ($t = 0.5$ s).

The regularity is found within a few repetitions. The brain's ability to find regular patterns can be reflected by the quickly 'adjusted' neural activities that reach another steady state within a few repetitions of the stimuli. Such *regularity representation* by neural activities can be observed in the adapted E populations mentioned above, as well as the gradually increased MEG amplitude (Figures 5.12 and 5.13). In the simulation where the three plasticity terms a, p^{EE}, and p^{IE} were considered, the E and I populations showed higher activities in REG than in RAND condition (Figure 5.16a and b). With a focus of the RAND-REG condition (left column of Figure 5.16), we can see the gradually increased E and I activities at the second repetition (around $t = 6$ s, left column in Figure 5.16b and c), which resembled the increased RMS of MEG responses shown in Figure 5.13b.

I further showed that the gradually increased activities (i.e., regularity representation) were due to the change in network connections through short-term plasticity (i.e., regularity encoding). In the simulation, the network connections were probed by the asymmetry indices AI^{EE} and AI^{IE} (Figure 5.16d), where the network connections started to form 'preference' to the input pattern (reflected by the increasing AIs). The preference was formed by the two plasticity terms p^{EE} and p^{IE}. The term p_{kj}^{EE} strengthened w_{kj}^{EE} at the sequence order $j \rightarrow k$, leading to an increase in

AI^{EE} and the increased E activities to the next repetition. Similarly, The term p_{kj}^{IE} weakened w_{kj}^{IE}, and the term p_{jk}^{IE} strengthened w_{jk}^{IE}, leading to an increase in AI^{IE} and the overall increased I activities.

The increases in AI^{EE} and AI^{IE} were accumulated more effectively in REG than in RAND condition. This can be seen by comparing the AIs (during $t = 0$ to 5 s) in the left and right columns in Figure 5.16d. The increase in AIs then lead to the increase in the E and I activities, and the speed of increase depended on the length of regular sequence. In short, the regularity was first encoded implicitly in the connections and then was represented explicitly by the neural activities.

The regularity representation collapses quickly once a violation happens. The collapse of regularity representation is much faster than the buildup. As shown in the histogram of time lags in Figure 5.13b, the transition is earlier and sharper in REG-RAND than in RAND-REG condition. This observation makes sense because a single deviant tone pip is enough to violate the regular pattern. In the simulation, we can see clear drop in E and I activities at the switch ($t = 5.5$ s) in REG-RAND condition (right column in Figure 5.16b and c; especially the I activities in Bank 2), because the deviant stimulus could not successfully activate the nodes that were either suppressed (due to asymmetric P^{IE}) or less enhanced (due to asymmetric P^{EE}). The drop of E and I activities was followed by the drop in AIs (right column in Figure 5.16d).
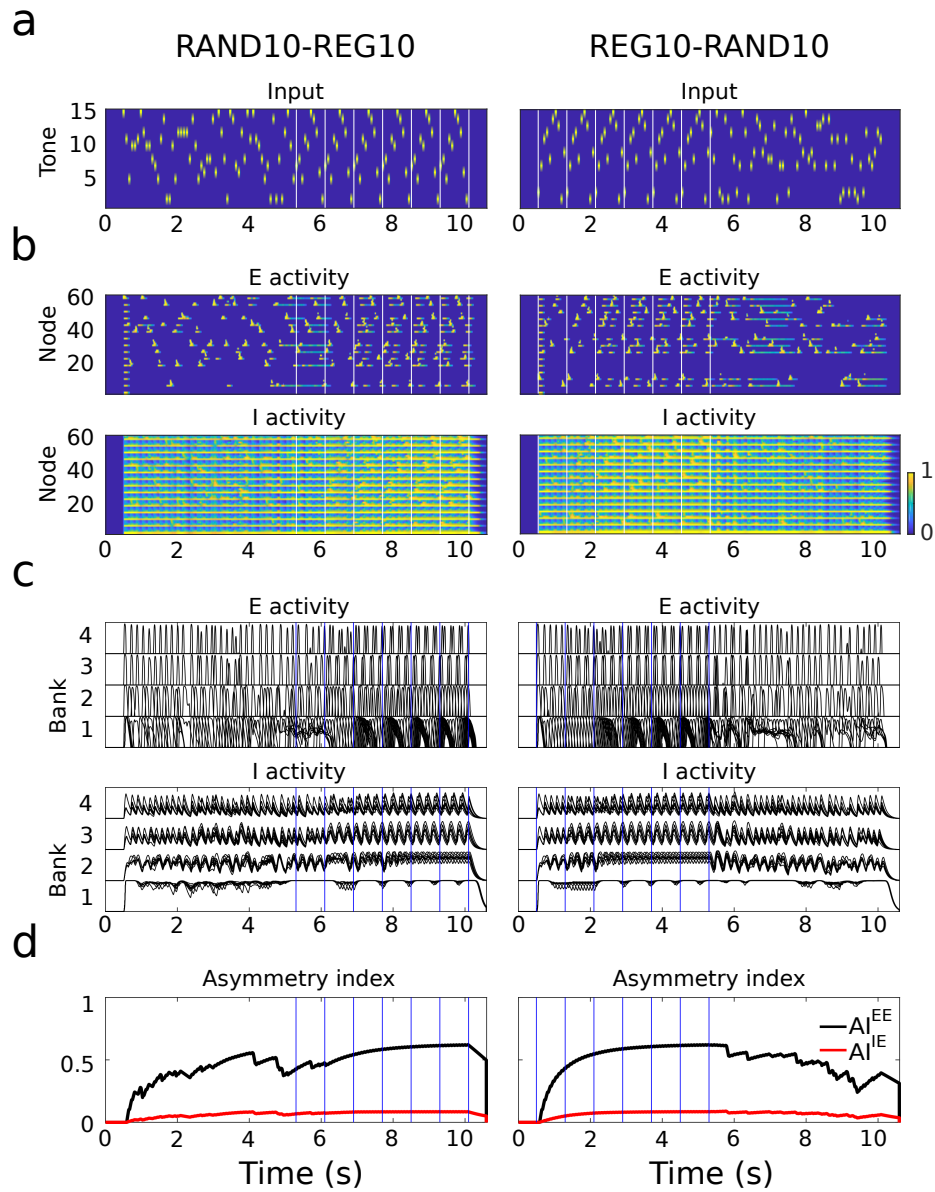
Figure 5.16: Excitatory and inhibitory activities in RAND-REG and REG-RAND conditions. (a) The external input sequences (duration= 80 ms, SOA= 80 ms, amplitude= 1 spikes/s, rise/fall time= 15 ms). The vertical white lines indicate the onset/offset of a regular sequence. (b) The activities (in firing rate, color coded) of the 60 E and I populations. (c) The 60 E and I activities are plotted separately in four resonant banks. Each bank comprises 15 nodes that receive the 15 tonal inputs. The nodes in the 1st bank have the weakest w_{jj}^{IE} and w_{jj}^{EI} and thus show slower dynamics than the nodes in other banks. (d) The asymmetry indices AI^{EE} (black curves) and AI^{IE} (red curves) are changing over time due to the plasticity terms p^{EE} and p^{IE}. Left column: condition RAND10-REG10. Right column: condition REG10-RAND10.

The *I* activities contribute to the level shift in MEG amplitude. The simulated MEG signals were generated along with the network activities shown in Figure 5.16. We can check the contributions of *E*-to-*E* and *I*-to-*E* currents to the simulated MEG signals by varying the respective parameters $r_{current1}$ and $r_{current2}$ described in Equation 3.33. Figure 5.17 shows the simulated MEG signals with $[r_{current1}, r_{current2}] = [1, 0]$, $[0, 1]$ and $[0.5, 0.5]$. The *E*-to-*E* currents contributed a lot to the On responses (Figure 5.17a). The *I*-to-*E* currents contributed to the level shift of MEG amplitude (Figure 5.17b), which resembled the prominent shift in RMS of MEG responses shown in Figure 5.13b. This suggests the importance of considering *I* activities in simulating MEG signals. The simulated MEG signals considering mixed *E* and *I* activities (Figure 5.13c) reproduced the MEG observation shown in Figure 6.3b[4].

---

[4]Except that the mismatch response and Off responses were not reproduced in this simulation, because I focused on regularity formation and did not include change detectors in the network.
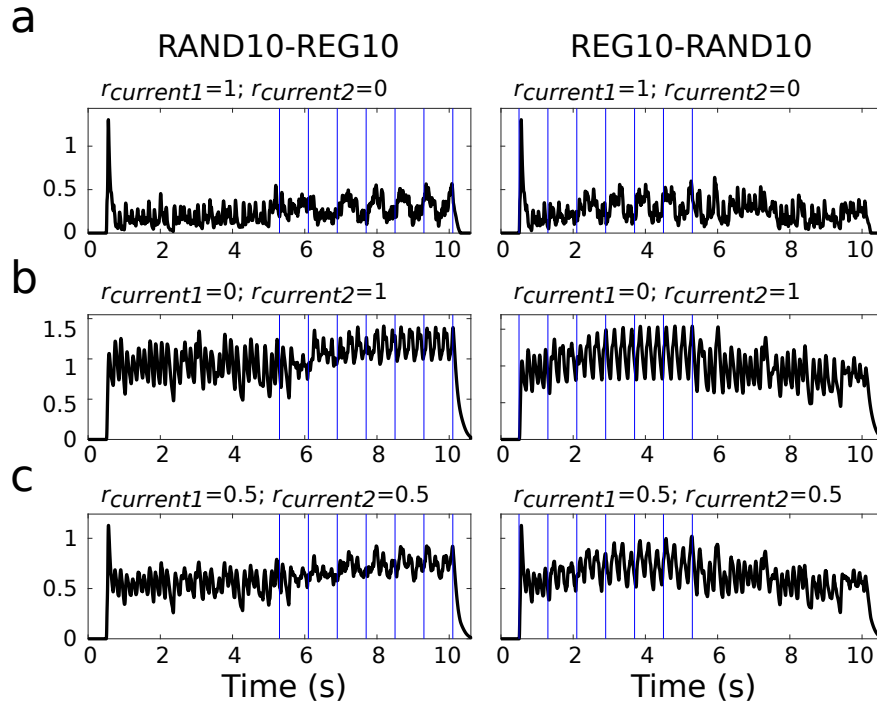
Figure 5.17: Simulated MEG signals in RAND-REG and REG-RAND conditions. (a) Contribution of the excitatory currents on the pyramidal cells (i.e., $r_{current1} = 1$ and $r_{current2} = 0$). (b) Contribution of the inhibitory currents on the pyramidal cells (i.e., $r_{current1} = 0$ and $r_{current2} = 1$). (c) Contribution of both the excitatory and inhibitory currents on the pyramidal cells (i.e., $r_{current1} = 0.5$ and $r_{current2} = 0.5$). The vertical blue lines indicate the onset/offset of a regular input sequence. Left column: condition RAND10-REG10. Right column: condition REG10-RAND10.

RMS amplitude correlates with input predictability. I have shown that the input stimuli alter the network connection through short-term plasticity (regularity encoding), which in turn affects the neural activities (regularity representation). Here I further assume that the correlation between RMS amplitude and input predictability (Figure 5.12) is due to short-term plasticity. Below I check the effect of the three plasticity terms a, p^{EE}, and p^{IE} on the amplitude of simulated MEG signals.

Figure 5.18a shows the simulated MEG signals in REGn-RANDn conditions (with alphabet size $n = 5, 10, 15$) where only the synaptic adaptation term a was considered. Figure 5.18b shows the mean MEG amplitude in REG ($t = 2$ to 4 s) and in RAND ($t = 7$ to 9 s) for alphabet size $n = 2, 3, ..., 15$. Shorter alphabet size n re-

sulted in slightly smaller mean MEG amplitude because the adaptation effect was stronger. In addition, the mean MEG amplitudes in RAND conditions (blue curve) were slightly smaller than in REG (red curve) because the synapses had on average shorter time to recover in RAND than in REG. Figure 5.18c shows the effective strength of w_{kj}^{EE} (black curve), w_{kj}^{IE} (red curve), and w_{jk}^{IE} (red dashed curve) at the last repetition (around $t = 4$ s) where the regular sequence included order $j \rightarrow k$. In this case the term a did not change too much of w_{kj}^{EE}.

Figure 5.18d shows the simulated MEG signals where the synaptic adaptation term a and the depression/facilitation term p^{IE} were considered. The MEG signals started to go a bit higher after the second repetition (indicated by the red, orange and green vertical lines). The MEG signals drop fast at the switch (indicated by the black vertical line). Figure 5.18e shows that the difference between REG and RAND was increased, because p^{IE} introduced less suppression on the regular order and more suppression on the other orders. Figure 5.18f shows that the effective strength of w_{kj}^{IE} and w_{jk}^{IE} were driven further away from initial values (red horizontal line) as the alphabet size decreased[5]. In this case, the simulation result fulfilled observed effect of regularity (i.e., group RMS: $REG_n > RAND_n$). However, the effect of alphabet size was not fulfilled (i.e., group RMS: $REG_n > REG_{n+1}$; $RAND_n > RAND_{n+1}$) because terms a and p^{IE} only induced adaptation and lateral inhibition among the E populations.

Figure 5.18g shows the simulated MEG signals where the three terms a, p^{EE}, and p^{IE} were considered. The MEG signals started to go up after the second repetition (indicated by the red, orange, and green vertical lines), which resemble the rise of RMS amplitude at the 8, 14, and 21 tones shown in Figure 5.12a. In Figure 5.18h, we see the simulation fulfilled the observed effect of regularity (i.e., group RMS: $REG_n > RAND_n$) as well as the effect of alphabet size (i.e., group RMS: $REG_n > REG_{n+1}$; $RAND_n > RAND_{n+1}$). Figure 5.18i shows that p_{kj}^{EE} enhanced a lot on w_{kj}^{EE} (black curve) as the alphabet size decreased.

The contributions of three short-term plasticity terms a, p^{EE} and p^{IE} for sequence regularity formation are summarized below.

1. The term a_{kj} ($k = 1, 2, ...N$) reduces the continuation of E_j activity, which contributes to the On responses in the MEG signals. It stabilizes the network

---

[5]alphabet size $n = 2$ was a special case because the effect of p_{kj}^{IE} and p_{jk}^{IE} canceled each other

from getting overwhelmed by the synchronous firing among the E popula-
tions.

2. The term p_{kj}^{EE} ($k = 1, 2, ...N$; $k \neq j$) promotes the activity of population E_k
   in the regular sequence order $j \rightarrow k$. This accounts for the effect of alphabet
   size (i.e., group RMS: $REG_n > REG_{n+1}$; $RAND_n > RAND_{n+1}$).

3. The term p_{kj}^{IE} ($k = 1, 2, ...N$; $k \neq j$) alters the 'impedance' of network con-
   nections by either reducing I_k activity in regular order $j \rightarrow k$, or increasing I_k
   activity in regular order $k \rightarrow j$. This accounts for the effect of regularity (i.e.,
   MEG amplitudes: group RMS: $REG_n > RAND_n$). The increased I activities
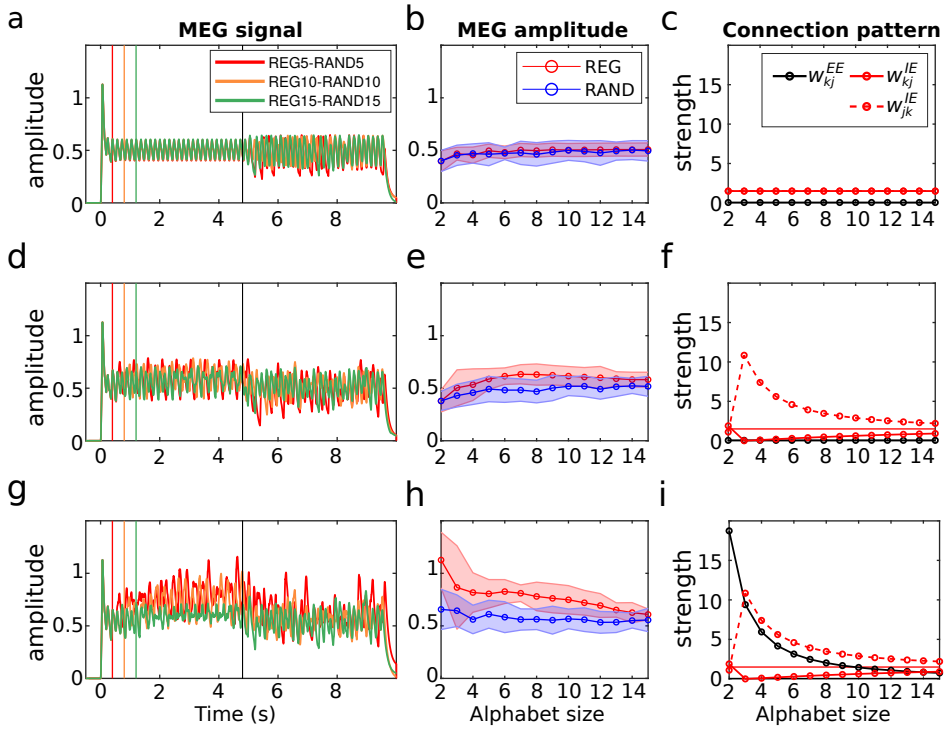   account for the prominent increased RMS amplitude in REG conditions.

Figure 5.18: The contribution of the three types of plasticity. (a) The simulated MEG signals in REG_n-$RAND_n$ condition ($n = 5, 10, 15$) when only the plasticity term a was considered. The red, orange and green vertical lines indicate respectively the start of the second repetition of sequence with alphabet size $n = 5$, 10, and 15. The black vertical line indicate the time of the switch from REG to RAND. (b) The mean and standard deviation of MEG amplitude during the last second in REG and in RAND. (c) The connections w_{kj}^{EE}, w_{kj}^{IE}, and w_{jk}^{IE} between Node j and k at the last repetition in REG, where the two nodes were activated in the order $j \rightarrow k$. (d-f) The plots where synaptic adaptation on W^{EE} and the sequence-order-dependent short-term plasticity on W^{IE} are considered. (g-i) The plots where synaptic adaptation on W^{EE} and the sequence-order-dependent short-term plasticity on W^{EE} and W^{IE} are considered.

5.4   Mechanism underlying regularity formation

In simulations III and IV, I used similar approaches to implement *regularity formation* where the principle is to represent a regular feature (e.g., regular tone, loudness, duration, periodicity, sequence, or abstract rule) in the stable neural activities that collapse in response to a violation. The approaches I used have three important ingredients in common:

1. **Place coding.** The place coding of the periodicity and tones were based on the tonotopic and periodotopic maps in the auditory cortex found in fMRI studies. It is crucial that features are place-coded so that the short-term plasticity can work on it (e.g., to connect neighboring nodes for sustained resonance, or to alter connection strength in favor of a sequence order). The current models in the simulations would not work if the input SOA were rate-coded by a single node.

2. **Large time constant τ_i.** The time constant τ_i of the I populations was set larger for regularity formation than for change detection. As illustrated in Figure 5.10, a large τ_i can increase the capacity of memory trace. Also, a large τ_i smoothens the neural activities, which stabilizes the change in connection strength due to short-term plasticity.

3. **Short-term plasticity.** To encode the regular features in the connection patterns at the timescale of hundreds of milliseconds, short-term plasticity seems to be a better candidate than long-term plasticity (e.g., spike timing dependent plasticity). Simulations III and IV have demonstrated how the regular feature can be encoded in the connections and represented by the stable neural activities.

The three ingredients are biologically plausible. In addition, the approach that utilizes place coding and short-term plasticity for regularity formation is economical, because the network dynamics can then have multiple stable points with the flexibility to reorganize and with a sufficient resolution of features[6], which is more practical than to design a network with a number of stable points to represent a single feature in a pseudo-continuous fashion.

The concept of regularity formation may have a similar counterpart in the other models of deviance detection. The short-term plasticity in the phase of *regularity encoding* was a generalized version of the synaptic adaptation used in the adaptation model. The *regularity representation* by neural activities can be analogous to the prediction signals in the prediction model. However, there is a clear difference. In the prediction model, the prediction is modified after a error signal is generated. In the proposed generic deviance detection principle, the regularity representation collapses before the change response is generated.

---

[6]The change responses can be viewed as transient trajectories during the switch between two stable points, so the distance between two stable points means the magnitude of change. This will be discussed in Chapter 6

More experimental evidence is required to refine this approach for regularity for-
mation. The difficulty is that the various forms of short-term plasticity of different
types of synapse have not been fully explored, and different combinations of the
plasticity rules may all give rise to acceptable simulation results. Even so, record-
ing the activities of excitatory and inhibitory neurons shall provide nice clues for
the modeling study.

Chapter 6

Deviance detection

"Before I had studied Zen for thirty years, I saw mountains as mountains, and rivers as rivers. When I arrived at a more intimate knowledge, I came to the point where I saw that mountains are not mountains, and rivers are not rivers. But now that I have got its very substance, I am at rest. For it's just that I see mountains once again as mountains, and rivers once again as rivers."

– Qingyuan Weixin

6.1 Overview

Deviance detection is a function that detects unexpected events. The proposed *generic deviance detection principle* suggests that unexpected events cause sudden changes in the representation of regularity, and these changes trigger the neighboring change detectors. In Chapter 4, I focused on the change detection and the underlying wiring patterns. In Chapter 5, I focused on the regularity formation and the neural representation of regularity. In this chapter, I come back to the generation of deviance-related responses, with a focus on the interaction between regularity formation and change detection.

I use the OSR and the sequence MMN as examples of the deviance-related responses. The elicitation of OSR requires the expectation to a constant periodicity of the repetitive stimuli, where the expectation can be encoded and represented by a network used in Section 5.2.2. Likewise, the elicitation of sequence MMN

requires the expectation to regular input sequence, where the expectation can be encoded and represented by a network used in Section 5.3.2. For demonstrations, simulation V showed that the OSR can be regarded as a change response to the cessation of constant periodicity. Simulation VI showed that the sequence MMN can be regarded as a change response to the switch in sequence regularity, where the change response can be a mixture of an On response to the deviant and an Off response to the cessation of regularity. At the end of this chapter, I provide my viewpoints on the confusions and conflicts (mentioned in Section 2.4) regarding the mechanism underlying deviance detection.

6.2   Omitted-stimulus response

6.2.1   Experimental observations

The omitted-stimulus response (OSR) differentiates itself from the Off response by its property of *temporal expectation*. The OSR possesses many properties[1] as observed in an MEG study [Andreou et al., 2015] shown in Figure 6.1:

1. The peak latencies of OSR are proportional to the stimulus onset asynchrony (SOA) of the repetitive stimuli. In order words, the OSR rises after a constant delay after the due-time, where the due-time means the time when the omitted stimulus was expected.

2. There are usually pre-activated peaks before the due-time. The pre-activated peaks rise before the OSRs, often with smaller amplitudes. The latencies of the pre-activated peaks are proportional to the SOAs, which reflect the temporal expectation.

3. There are different MEG waveforms during the repetitive stimuli of different SOAs. This reflects the resonant property of brain networks collaboratively encoding the periodicity. The adaptive mechanism can be seen in the gradually down-going curves and the subtle tuning in the waveforms.

These properties can be accounted for by the generic deviance detection principle. Below in the simulation (Section 6.2.2), I demonstrated that the OSR can be

---

[1]These properties are ubiquitous. See other studies mentioned in Section 2.2.2.

regarded as a change response (or an Off response) to the cessation of a stable periodicity.
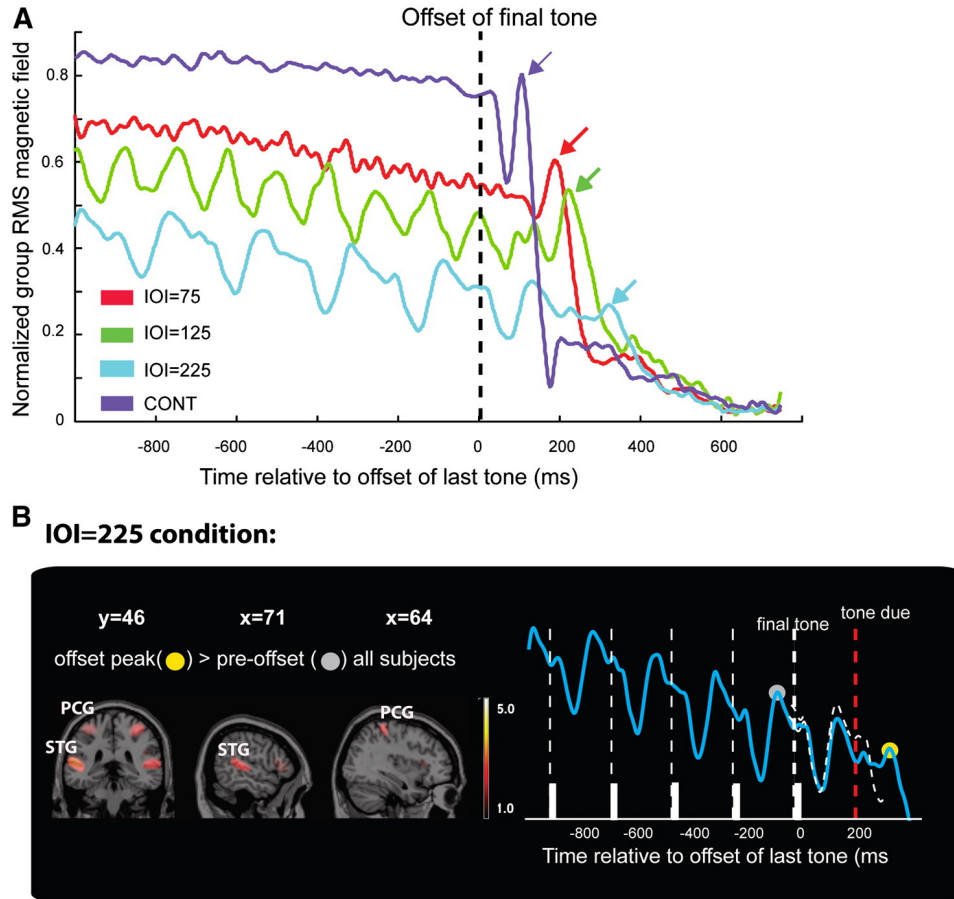

Figure 6.1: Omitted-stimulus response (MEG). (A) Group RMS (RMS of individual subject RMSs across 40 channels) time-locked to the offset of the last tone ($t = 0$ ms). The stimulus input contained either a long pure-tone stimulus (CONT) or sequences of 25 ms tone burst with inter-onset-intervals (IOI) of 75, 125, and 225 ms. The arrows indicate the response peaks (mean latencies: CONT = 95 ms, IOI75 = 195 ms, IOI125 = 227 ms, and IOI225 = 317 ms). (B) A focus on the IOI225 condition. The offset peak (yellow circle) rose after the tue-time (red dashed line). There was another peak rising before the due-time, which resembled the pre-offset peak (gray circle). The significant difference (offset peak > pre-offset) was sourcelocalized to the bilateral superior temporal gyrus (STG) and post-central gyrus (PCG). (Figure copied from [Andreou et al., 2015])

6.2.2 Simulation V: omitted-stimulus response

This simulation demonstrated (1) the *temporal expectation* illustrated in Figure 6.2a, where the peak latency of the simulated MEG signal is equal to the SOA plus

a constant delay d, (2) the pre-activated peak before OSR, and (3) the different waveforms across SOAs.

Simulation settings

Network and input. The network comprised $N = 21$ nodes (Figure 6.2b), where Nodes 1 to 20 formed a *resonant bank* (or a bank of oscillators. See also Figure 5.4a) and Node 21 was a change detector connected to the resonant bank with re-scaled[2] inter-node connections selected from the W solutions of Dec-Off type (see also Section 4.2.2). The resonant bank comprised nodes with different resonance frequencies [May and Tiitinen, 2001, Large et al., 2010]. The use of a resonant bank is based on the observation that the auditory cortex shows a spatial representation of both frequency and periodicity [Langner et al., 2009, Barton et al., 2012, Baumann et al., 2011]. Previously, the resonance frequency of a node was controlled by its intra-node connections and a proper background input for a better linear working range (Figure 5.5 in Chapter 5). In this simulation, the resonance frequency of a node was controlled by its time constants τ_e and τ_i for simplicity, where the background inputs to every node could then be set the same. The external inputs were fed to the resonant bank. An input trial was a prolonged stimulus (duration= 7000 ms, amplitude= 1 spikes/s, rise/fall time= 10 ms) or repetitive stimuli (SOA= 75 to 250 ms, duration= 50 ms, amplitude= 1 spikes/s, rise/fall time= 10 ms). Detailed configurations were specified in Table 6.1[3].

Short-term plasticity. Short-term plasticity was applied to the inter-node connections W^{EE} and W^{EI} in the resonant bank[4] as a possible solution for the regularity formation of input periodicity. The general form of plasticity term p_{kj} (j≠k) described in Equation 3.32 is rephrased into below Equations 6.1 and 6.2, where p_{kj}^{EE} strengthen the efficacies w_{kj}^{EE}, and p_{kj}^{EI} strengthen w_{kj}^{EI} [5]. The plasticity term p_{kj}^{EE} increases if the covariance $Cov_{j,k,\Delta t}(t)$ between $m_j^E(t)$ and $m_k^E(t)$ from time $t - \Delta t$ to t is positive, and otherwise decreases gradually back to zero (Equations 6.1). Similarly, the plasticity term p_{kj}^{EI} increases if $Cov_{j,k,\Delta t}(t)$ is negative, and otherwise

---

[2]The inter-node connections W need to be re-scaled according to the number of nodes in the resonant bank, because the W solutions are based on the two-node network.

[3]Other parameters were same as the default values in Table 3.1.

[4]Applying short-term plasticity on W^{EE} and W^{IE} also works in this example.

[5]Synaptic adaptation was considered in simulation III (Section 5.2.2) but not considered in this simulation. It turned out that the plasticity term p_{kj}^{EI} had a similar effect as the synaptic adaptation term a_{kj} in this case.

Table 6.1: Configurations for simulation V: omitted-stimulus response

Par.	Value (Node 1-20)	Value (Node 21)	Unit
Impulse response function			
τ_e	4 to 13	20	ms
τ_i	$2\tau_e$	40	ms
H_e	$3.25 \times 10/\tau_e$	$3.25 \times 10/\tau_e$	mV
H_i	$22 \times 20/\tau_i$	$22 \times 20/\tau_i$	mV
External connections			
w^{EX}	220×0.3	0	1
w^{IX}	220×0.15	0	1

Par.	Value	Unit
Intra-node connections ($j = 1, 2, ..., 20; k = 21$)		
$[w_{kj}^{EE}, w_{kj}^{IE}, w_{kj}^{EI}, w_{kj}^{II}]$	$135 \times 10^{-3} \times [12, 6, 6, 6]$	1
$[w_{jk}^{EE}, w_{jk}^{IE}, w_{jk}^{EI}, w_{jk}^{II}]$	$135 \times 10^{-2} \times [6, 6, 12, 12]$	1

decreases gradually back to zero (Equations 6.2). The weight masks g_{kj}^{EE} and g_{kj}^{EI} consider the effectiveness of plasticity as a function of the distance between nodes j and k. For 20 nodes in the resonant bank, the weight masks follow the Gaussian function $exp\left(-(j-k)^2/2\sigma^2\right)$, where σ was set to 4 and 8 for g_{kj}^{EE} and g_{kj}^{EI}, respectively.

$$\dot{p}_{kj}^{EE}(t) = \frac{-p_{kj}^{EE}(t)}{\tau_p^{EE}} + \kappa_p^{EE} \cdot g_{kj}^{EE} \cdot \left[Cov_{j,k,\Delta t}(t)\right]_+ \qquad (6.1)$$

$$\dot{p}_{kj}^{EI}(t) = \frac{-p_{kj}^{EI}(t)}{\tau_p^{EI}} + \kappa_p^{EI} \cdot g_{kj}^{EI} \cdot \left[-Cov_{j,k,\Delta t}(t)\right]_+ \qquad (6.2)$$

Simulated MEG. The simulated MEG signals $R(t)$ were generated according to Equation 3.6. In order to highlight the contribution of node 21 (change detector), the node weight r_{21} was set three times larger than the other weights r_{1-20}. The contribution of positive and negative ion currents were set equal: $r_{current1} = r_{current2} = 1$.
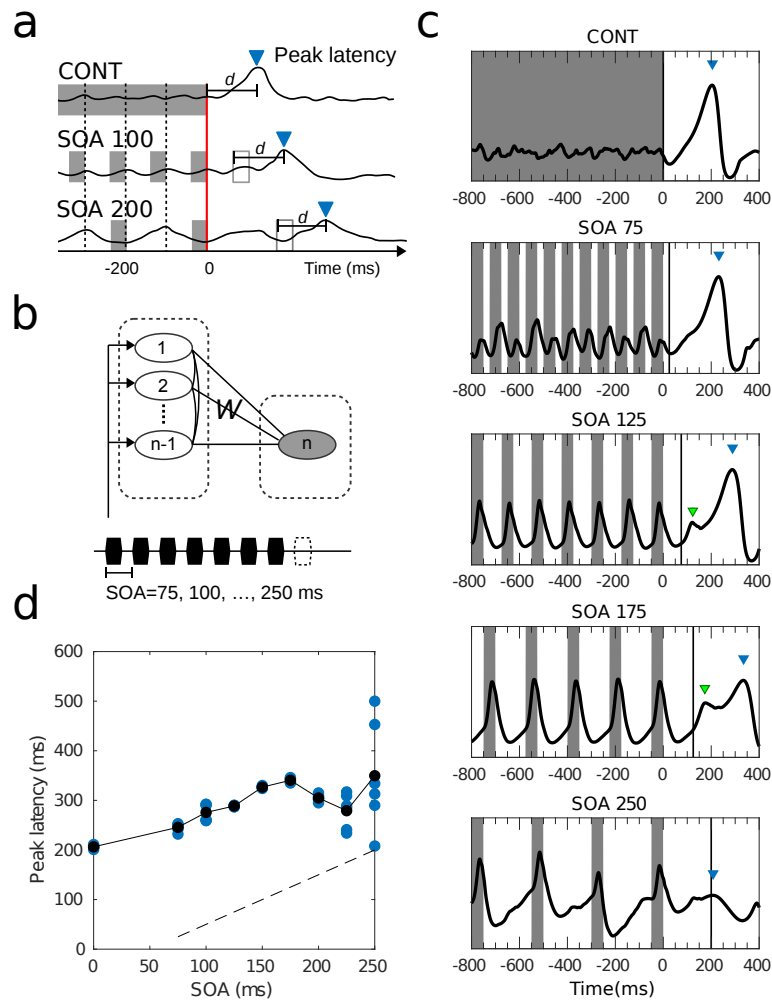
Figure 6.2: Omitted-stimulus response (OSR). (a) Illustrative responses that show temporal expectation. The peak latencies should be linear to the SOA if the offsets of stimuli are aligned (red line), or a constant d if the due-times are aligned (empty rectangles). See, for example, Figure 2 in [Andreou et al., 2015]. (b) In the simulation settings, the R nodes (left column) are simply implemented with different time constants τ_e and τ_i, leading to different resonance frequencies. A prolonged stimulus or periodic stimuli are fed to these R nodes. (c) The simulated MEG signals (black curves) rise after the due-time (black vertical lines) and show different peak latencies and peak amplitudes (marked with blue triangles). The small peaks (marked with green triangles) reflect the momentum of the bank of oscillators. (d) Simulated peak latencies are linear to the SOA. Simulations are run for several trials for each SOA where the offset time is changed. The peak latencies in each simulation trial (blue dots) under the same SOA can be different, which depends on the network stability during the stimulus and the offset time. Black dots are the mean peak latencies. The peak latencies show an approximately constant delay with respect to due-time (time of predictable omission, dashed line) when the SOAs are below 200 ms. The peak latencies become unstable across trials when SOAs are above 200 ms. In other words, the temporal expectation is preserved in this network for SOAs smaller than 200 ms.

Simulation results

Figure 6.2c shows the simulated MEG signals under prolonged (CONST) and periodic stimuli (SOAs: 75, 125, 175, and 250 ms; stimulus duration: 50 ms). As the input SOA increased, the latency of the OSR peaks (blue triangles) increased, and the amplitude decreased. This is in line with MEG observations [Andreou et al., 2015]. The small pre-activated peaks before the OSR (particularly clear for SOA 125 and 175; green triangles) were located at the time of the omitted stimulus, which resembles the expected evoked potential before the OSR (e.g., Figure 7B in [Bullock et al., 1990]). The resonant bank with short-term plasticity enabled the tolerance of a range of input SOAs, the sustained resonance that causes the pre-activated peak, and the different resonant waveforms during the repetitive stimuli.

Figure 6.2d shows that the network is able to respond with the correct timing (i.e., a constant delay after due-time) if the SOA is within 150 ms. The peak latencies become unstable for SOAs larger than 200 ms. This limitation is due to the limit of resonance frequencies in the resonant bank. As can be seen in Figure 6.2c, the simulated MEG data for SOA 250 is not as stable compared to the faster SOAs.

In this example, I demonstrate that the cortical OSR can reflect a detection mechanism upon the stable representation of periodicity. The sustained resonance was crucial for temporal expectation. This is in line with the observation that the auditory brainstem does not generate overt OSRs [Lehmann et al., 2016], likely because sustained resonance has not happened at that stage. Source analysis, as well as fMRI, showed that the OSR (more specifically, the fast OSR [Karamürsel and Bullock, 2000]) is localized in the auditory cortex [Raij et al., 1997, Mustovic et al., 2003, Yamashiro et al., 2009, Andreou et al., 2015], suggesting that the auditory cortex has the capacity to represent a certain range of periodicity locally (e.g. under 200 ms). However, I have not yet fully investigated the neural mechanism underlying temporal expectancy. The resonant bank, which only assumes heterogeneity across neural populations, is so far a good candidate for implementation.

6.3   Mismatch negativity

6.3.1   Experimental observations

The responses in a roving paradigm reveal the progress of regularity formation and change detection, and thus are useful for demonstrating the generic deviance detection principle. In Figure 6.4a, an MEG study shows how the human brain responds to the switch between regular and random complex acoustic patterns [Barascud et al., 2016]. There are On and Off responses at the onsets and offsets of the stimulus sequence. An MMN response is elicited by the transition from regular to random sequences (REG-RAND), while there is only a gradually rising root mean square (RMS) amplitude the other way around (RAND-REG). Also, the RMS amplitude is higher during regular sequences compared to random sequences.
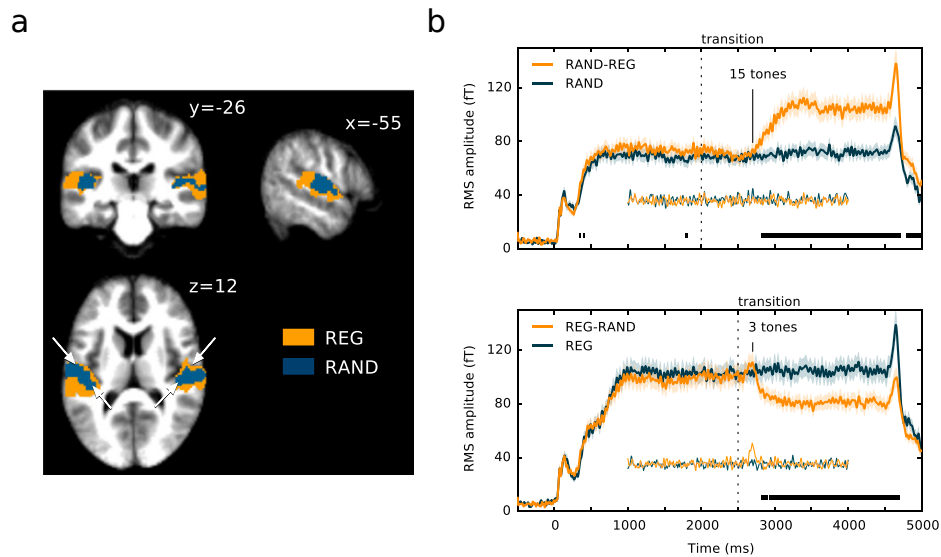

Figure 6.3: Brain activities during regular (REG) and random (RAND) sequences. (a) fMRI group activation for REG and RAND conditions. The white arrows indicate the midline of Heschl's gyrus. (b) MEG responses (RMS of individual RMSs) to RAND-REG and REG-RAND conditions (yellow curves) as well as the RAND and REG conditions (blue curves). See also individual RMSs in Figure 5.13. The horizontal black lines indicate significant differences between the blue and yellow curves. The vertical dashed lines indicate the time of transition. The responses after high-pass filtering are shown underneath the main curves. A mismatch response can be seen in the filtered response in REG-RAND condition. (Figure copied from [Barascud et al., 2016])

6.3.2   Simulation VI: sequence MMN

This simulation demonstrated (1) On and Off responses at the onset and offset of stimulus sequences, (2) MMN response to the transition from regular to random sequences, and (3) different RMS amplitudes during REG and RAND presentations.

Simulation settings

Network and input. The network comprised three nodes, where two nodes were responsible for *regularity representation*, and one node served as a change detector (Figure 6.4a). Nodes 1 and 2 received the respective inputs RAND and REG (prolonged stimuli; amplitude= 1 spikes/s; rise/fall time= 10 ms) representing the identities of the random and regular sequences. The duration of the inputs RAND (2 s) and REG (2.5 s) were set to match the experimental results in Figure 6.3b. The two inputs and Nodes 1,2 were not meant to model the physical stimuli and the corresponding cortical areas[6], but to capture the fMRI observation showing the different portions of activation for RAND and REG conditions (Figure 6.3a).

Short-term plasticity. The inter-node connections between nodes 1,2 and node 3 were selected from the *W* solutions of Inc-OnOff type, and the inter-node connections between nodes 1 and 2 were tuned manually to match the observed RMS waveforms in Figure 6.3b. No short-term plasticity was used in this simulation because the actual process of regularity formation was not considered. Detailed configurations were listed in Table 6.2.

Simulated MEG. The simulated MEG signals were generated according to Section 3.6. In order to highlight the contribution of Node 3 (change detector), the node weight r_3 was set six times larger than the other weights $r_{1,2}$. The contribution of positive and negative ion currents ($r_{current1}$ and $r_{current2}$) were both set to 1.

---

[6]The network of 60 nodes for *regularity formation* in simulation IV in Section 5.3.2 was meant to model the the physical stimuli and the corresponding cortical areas. The regularity representation by Nodes 1 and 2 is simply a result of regularity formation

Table 6.2: Configurations for simulation VI: sequence MMN

| Par.	Value	Unit
Impulse response function		
τ_e	10	ms
τ_i	20	ms
H_e	3.25	mV
H_i	22	mV
Sigmoid function		
e_0	2.5	spikes/s
r	0.56	1/mV
v_0	6	mV
Intra-node connections		
$[w^{EE}, w^{IE}, w^{EI}, w^{II}]$	$135 \times [0.8, 0.6, 0.2, 0.05]$	1
Background inputs		
B	220×0.5	spikes/s
External inputs and connections ($q = $ RAND, $k = 1$; $q = $ REG, $k = 2$)		
Amplitude of $x_q(t)$	1	spikes/s
w_{kq}^{EX}	220×0.3	1
w_{kq}^{IX}	220×0.15	1
Inter-node connections ($j = 1, 2$; $k = 3$)		
$[w_{kj}^{EE}, w_{kj}^{IE}, w_{kj}^{EI}, w_{kj}^{II}]$	$135 \times [0.1, 0.4, 0.1, 0.2]$	1
$[w_{jk}^{EE}, w_{jk}^{IE}, w_{jk}^{EI}, w_{jk}^{II}]$	$135 \times [0.1, 0.1, 0.2, 0.1]$	1
Inter-node connections ($j = 1$; $k = 2$)		
$[w_{kj}^{EE}, w_{kj}^{IE}, w_{kj}^{EI}, w_{kj}^{II}]$	$135 \times [0.4, 0.4, 0.2, 0.2]$	1
$[w_{jk}^{EE}, w_{jk}^{IE}, w_{jk}^{EI}, w_{jk}^{II}]$	$135 \times [0, 0, 0.25, 0.25]$	1

a

b
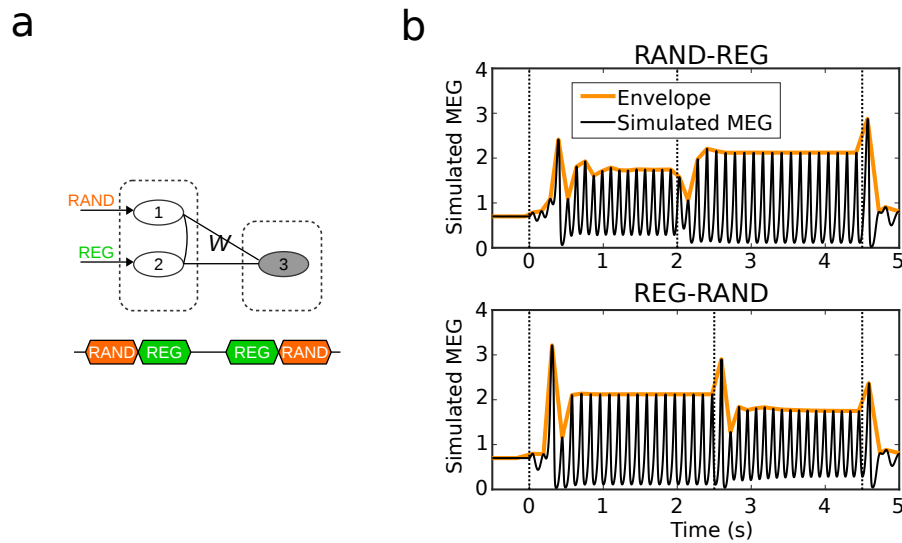
Figure 6.4: The network and the simulated MEG signals for sequence MMN. (a) In the simulation settings, a three-node network is used for mimicking the observation in Figure 6.3b. The R nodes (Nodes 1 and 2) receive the inputs representing RAND (orange prolonged stimuli) and REG (green prolonged stimuli), respectively. The inter-node connections W between the C node (Node 3) and the R nodes were selected from the W solutions such that the C node shows Inc-OnOff responses to the inputs. The connections between Nodes 1 and 2 were tuned manually so that the two nodes inhibit each other with different strength. This resulted in the different MEG levels during RAND and REG conditions. (b) Simulated MEG signals of the three-node network. The vertical dashed lines indicate the onsets/offsets/switches of inputs RAND and REG.

Simulation results

Figure 6.4b shows the simulated MEG signals of the three-node network in the RAND-REG and REG-RAND conditions. There were On and Off responses at the onsets and offsets of the inputs (around $t = 0.5$ s and 4.5s). There was a mismatch response at the switch from REG to RAND (around $t = 2.5$ s, lower panel), but not at the switch from RAND to REG (around $t = 2$ s, upper panel). Figures 6.5 and 6.6 show the firing rates and PSPs of the three nodes, in which we can see how the inter-node connections affected the interaction between the nodes. Figure 6.7 showed the phase portraits of Node 3, in which we can understand why there was no mismatch response at the switch from RAND to REG.

In Figures 6.5 and 6.6, the onsets of RAND-REG and REG-RAND trigger the On responses in Node 3, the *transient disinhibition* as described in Section 4.3.1. The

offsets of RAND-REG and REG-RAND trigger the Off response in Node 3. The Off responses are highlighted by the rectangles $P1$ and $P2$, where we can see the inhibited I_3 PSP before the *release from long-lasting disinhibition* as described in Section 4.3.2. Node 3 shows increased activities (especially the firing rate $m_3^E(t)$), because the inter-node connections between Nodes 1,2 and Node 3 belongs to the W solutions of Inc type. The activities of Node 2 during REG were higher than those of Node 1 during RAND, because I_2 strongly inhibit E_1 and I_1 (highlighted by the green rectangles) so Node 1 has no chance to inhibit Node 2 during REG. On the other hand, Node 2 slightly inhibits Node 1 during RAND because I_2 was still active (orange rectangles). In short, the unbalanced inhibition between Nodes 1 and 2 resulted in the different levels of network activity in REG and RAND conditions.

The absence of a mismatch response at the switch from RAND to REG can be better explained by checking the transitions in the phase portraits. In Figure 6.7a, b, and d, the transitions (black curves) resembles the ones in the offset detection in Figure 4.6b and d, where the clockwise trajectories flip up, show a transient peak, and become counter-clockwise. In Figure 6.7c, the transition from RAND (orange trajectory) to REG (green trajectory) goes in the other direction, where both $v_3^E(t)$ and $v_3^I(t)$ are pulled more negatively, and there is no chance for a 'release from disinhibition.' This explains why Node 3 does not responds to the switch from RAND to REG, even though it responded to the RAND offset and the REG onset. In other words, the response of a change detector is not intuitively a superposition of its Off response to one event and its On response to another. Rather, the response depends on the transition between the two steady states of the change detector[7].

I did not use more realistic settings. The rise/fall time of the two inputs could have been set differently, because it is reasonable to set a longer rise time for the REG stimulus input because it takes some time (at least a sequence length) to form regularity representation. This could directly explains why there was no mismatch response in the RAND-REG transition. Moreover, the intensity of the two stimulus inputs could have been reasonably set differently because the status of neural populations under REG and RAND can be dramatically different. This could directly explain the level changes in RMS amplitude. In this simulation, I set the inputs REG and RAND identical, in an attempt to highlight the effect of inter-node con-

---

[7]This is different from the case of multiple change detectors, where the overall mismatch response can be the sum of On responses in some nodes and Off responses in other nodes

nections W on the reshaping of the network activity. Again, this simulation did not focus on the details of regularity formation. To understand how the REG sequence causes higher RMS amplitude, I assumed short-term plasticity on W^{EE} and W^{IE} is required in the lower-level neural populations at the stage of regularity formation (See simulation IV in Section 5.3.2).
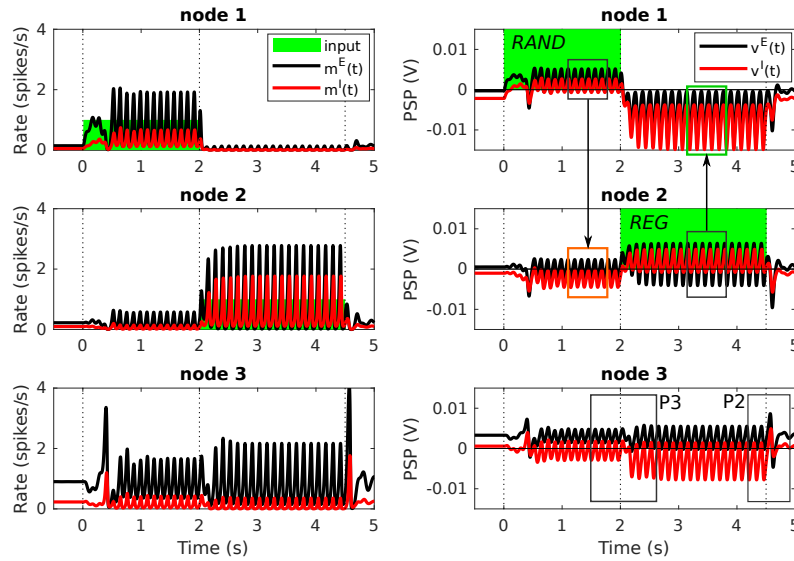

Figure 6.5: Activities of the three-node network (RAND-REG). The left column shows the firing rates of the three nodes, and the right column shows the PSPs. The activities (firing rates and PSPs) of the E and I populations are shown in the black and red curves, respectively. The inputs (RAND and REG) are shown in the green areas. The vertical dashed lines indicate the onsets/offsets/switches of the REG and RAND sequences. The black rectangles highlight the interaction between the nodes. See details in simulation results. The trajectories P2 and P3 are shown in Figure 6.7b and c.
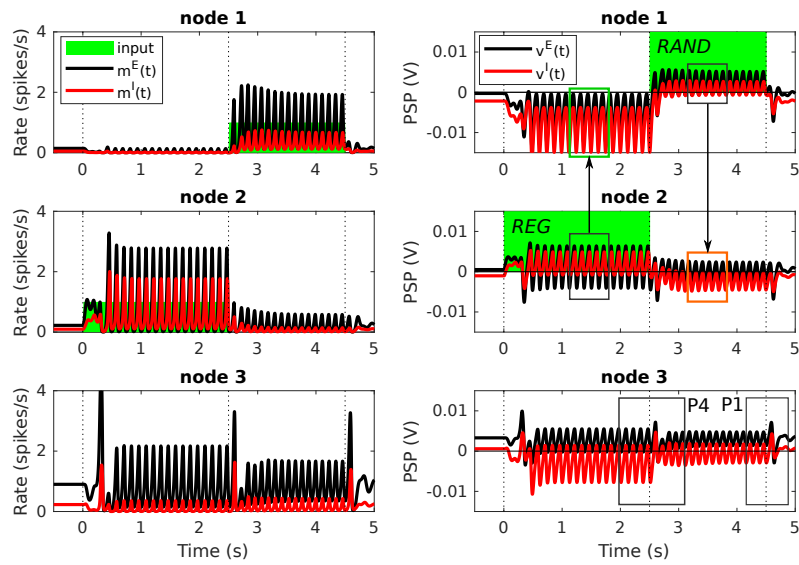
Figure 6.6: Activities of the three-node network (REG-RAND). Same settings as in Figure 6.5. See details in simulation results. The trajectories P1 and P4 are shown in Figure 6.7a and d.
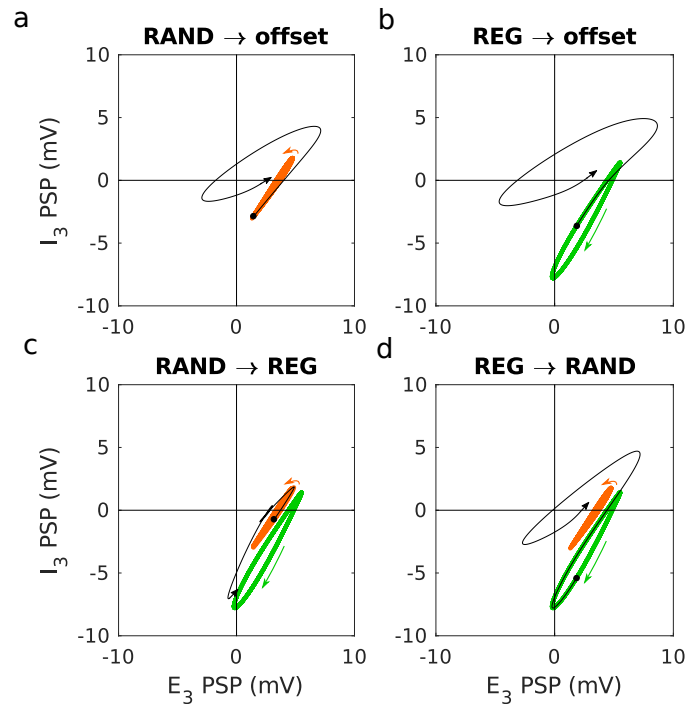
Figure 6.7: Phase portraits of Node 3. The orange and green phase portraits represent the steady state during RAND and REG, respectively. Black dots indicate the time of the offset/switch. The black curves show the trajectories (200 ms) during the transition. (a) An Off response at the offset of RAND. i.e., the P1 in Figure 6.6. (b) An Off response at the offset of REG. i.e., the P2 in Figure 6.5. (c) No peak at the switch from RAND to REG. i.e., the P3 in Figure 6.5. (d) A mismatch response at the switch from REG to RAND. i.e., the P4 in Figure 6.6.

6.4   Mechanism underlying deviance detection

The OSR is not just sustained resonance. The OSR is differentiated from the Off response by its peak latencies that are proportional to the SOA in repetitive stimuli, reflecting the role of temporal expectancy. To maintain a short continuation of neural activity (i.e., sustained resonances) that preserves the periodicity of the repetitive stimuli, models that claim to account for the OSR utilize either an *adaptive approach* [Thivierge and Cisek, 2011] or *place coding approach* [May and Tiitinen, 2001]. However, sustained resonances alone cannot fulfil all observations in terms of peak amplitude and peak latency of the response. First, for the peak amplitude, the OSR cannot simply rely on the sustained resonance since the ampli-

tude of OSR can be stronger than the evoked response during entrainment [Horváth et al., 2010]. Second, for the peak latency, there should be a constant delay following the due-time after stimulus offset [Andreou et al., 2015, Schwartz and Berry II, 2008], but the sustained resonance rises exactly at the due-time. Therefore, even though the sustained resonance is time-locked to the subsequent stimulus, there seems to be additional neural circuits responsible for the extra delay in peak latency and the stronger peak amplitude than the evoked responses. In simulation III, I demonstrated that the simulated OSR solves the two issues mentioned above (Figure 6.2). The model suggests that the cortical OSR can be interpreted as a cortical Off response at the end of sustained resonance. The simulation results are also in line with the finding that there is a pre-activated response at the time of expected onset followed by a mismatch response [Andreou et al., 2015, Bendixen et al., 2009, SanMiguel et al., 2013].

The OSR is not a prediction signal. The *omission paradigm* is often used to differentiate the contribution of *adaptation* and *prediction* in MMN generation. This is based on the assumption that the OSR could not arise without a stimulus and the involvement of active prediction. Interestingly, the models based on either the adaptation or prediction hypotheses interpret the OSR as essentially different from the MMN that is triggered by the classic oddball paradigm. In the adaptation-based model, the OSR is regarded as a rebound response (i.e., sustained resonance) rather than a delayed N1 [May and Tiitinen, 2001, May and Tiitinen, 2010]. In the prediction-based model, the OSR is regarded as a pure prediction signal that originates from the memory unit rather than prediction error [Wacongne et al., 2012]. Both interpretations imply that the OSR is essentially different from the MMN because no additional NMDA current is generated. The problem is that neither the rebound response nor the prediction signal explain the two observations in terms of amplitude and latency mentioned above. As demonstrated in simulations III and IV (Figure 6.2 and 6.4), I suggest that the cortical OSR and MMN are essentially the same, both being the activity of change detectors.

The *cross-modal omission paradigm* is also used to emphasize the need for prediction. The brain can predict an upcoming event (e.g., a handclap sound) from the preceding events of another modality (e.g., a silent handclap video, or self-paced button press) and an OSR is triggered if an expected stimulus is omitted. In a motor-auditory (MA) paradigm, participants show OSRs when the sound, expected to be initiated by the self-paced button press, is omitted [SanMiguel et al., 2013].

In a visual-auditory (VA) paradigm, an OSR is elicited by occasionally omitting the sound that accompanied a handclap video [Stekelenburg and Vroomen, 2015]. To date, cross-modal OSRs have not been considered by computational models. How does the generic deviance detection principle view the OSRs in these cross-modal paradigms that seem to be bound to an active predicting process? Here I provide my viewpoint. First, the prediction is likely to be supported by the association between the cross-modal events (e.g., handclap video or button press, followed by a sound stimulus) that have to be paired or learned (e.g., by Hebbian learning) in advance via direct or indirect connections. The existence of association is reflected by the pre-activation at 40 to 80 ms in the auditory cortex elicited by a visual event [Stekelenburg and Vroomen, 2015] or by a motor event [SanMiguel et al., 2013, Stekelenburg and Vroomen, 2015]. In the MA paradigm there is no pre-activation in the auditory cortex in the random condition where the button press is followed by a randomly selected sound and there is also no OSR thereafter [SanMiguel et al., 2013]. This suggests 48 trials is not enough to associate the button press to all 48 sound samples. Second, due to the pre-activation in the auditory cortex, the MA and VA paradigms can then be regarded as classic oddball paradigms where the standard is a 'weak-strong' sound pair and the deviant is a 'weak-omission' sound pair. In this sense, the cross-modal omission paradigm resembles an 'intensity MMN' or 'duration MMN' paradigm rather than an omission paradigm. This analogy explains why OSRs are elicited in the VA and MA conditions but not in auditory-only conditions (like a classic omission paradigm) [Stekelenburg and Vroomen, 2015]. More specifically, the SOAs (average 1155 ms) in the paradigm are above the temporal window of integration (TWI) for temporal features such as periodicity, but still within the TWI for identity features such as intensity and duration. The analogy can be verified if the VA and MA conditions fail to elicit 'omission' responses when the SOAs are larger than TWI for identifying features. Based on this analogy, the deviance detection that takes place in the auditory cortex stands alone from the process of association. This would explain why the pre-activation does not differ when the chance of sound omission is 50% versus 12%, while the mismatch response following the pre-activation depends on the proportion of omission trials for both VA and MA conditions [Stekelenburg and Vroomen, 2015]. Association is less likely to be reduced by the 50% omissions, whereas deviance detection relies much more heavily on probability. Taken together, given the pre-activation via association and the analogy to the classic MMN paradigm, computational models that account for the classic MMN (e.g., either prediction-based or not) could poten-

tially also account for the mismatch responses in cross-modal omission paradigms. From the viewpoint of generic deviance detection principle, the process of deviance detection (including regularity formation and change detection) takes place locally in the auditory cortex, even in the case of cross-modal VA and MA paradigms.

Different processes in regularity formation, but the same mechanism in change detection. The generic deviance detection principle suggests that *change detection* may rely on a common neural mechanism (i.e., the local reciprocal wiring), while *regularity formation* may, depending on the level of abstraction, require different brain resources and time to collect relevant information.

There are a number of dissimilarities among deviance-related responses, which, as discussed below, are mainly due to differences in the process of regularity formation. I take the differences between cortical OSR and MMN as an example. In terms of the temporal window of integration (TWI), a pitch MMN can be elicited by traditional oddball paradigms even when the SOA is larger than 500 ms [Sams et al., 1993, Bartha-Doering et al., 2015], while the estimated length of TWI for cortical OSRs is much shorter (160-170 ms) [Yabe et al., 1998]. In terms of attention, it has been suggested that fast and slow periodic sequences elicit cortical OSRs by two different mechanisms: The fast OSR (periodicity > 5 Hz) is elicited automatically, while the slow OSR (periodicity < 2 Hz) requires the involvement of attention [Karamürsel and Bullock, 2000]. The slow OSR can be elicited at large SOAs such as 800 ms in [Halgren et al., 1995]; and 1000 to 2000 ms in [Busse and Woldorff, 2003]. The need for attention suggests that the cortical OSR and MMN are different processes [Ng and Penney, 2014]. In terms of required repetition, a successful elicitation of MMN needs only two or three repetitions for simple feature-repetition regularities [Bendixen et al., 2007, Cowan et al., 1993, Sussman et al., 2007, Winkler et al., 1996], while the cortical OSR requires up to 9 repetitions in a train for a successful elicitation [Horváth et al., 2010]. The above observations suggest different processes, related to the degree of difficulty in regularity formation, underlie cortical OSR and MMN.

There are also several similarities among the deviance-related responses that support the notion of a common mechanism for change detection. In terms of latency, the peak latencies of cortical On/Off responses, cortical OSRs, and the MMN all fall in the range of 100-200 ms [Andreou et al., 2015, Näätänen et al., 1989, Rinne et al., 2006, Yabe et al., 1998, Raij et al., 1997]. In terms of spatial distribution, the sources of cortical Off response and MMN are similar. As revealed in animal

studies, the sources of Off responses appear to be in the non-tonotopic area adjacent to the tonotopic area [Takahashi et al., 2004, Baba et al., 2016]. In dense mapping MMN studies, the pitch MMN was reported to be generated in the secondary auditory area (or spreading more widely over the core and belt areas). This is distinct from the sources of the P1 and N1, at the core areas (A1 and AAF) [Shiramatsu et al., 2013, Pincze et al., 2001]. Cortical responses to the onset, offset, and pitch change of a continuous stimulus all share similar topography and temporal profiles, as suggested in several EEG/MEG studies [Yamashiro et al., 2011, Yamashiro et al., 2009, Nishihara et al., 2011]. Deviance-related responses also show similarities in their dependency on several factors regarding the regularities (e.g., probability of deviant, randomness in SOA, number of repetitions, effect of the NMDA-r antagonists) and the deviance magnitude (e.g., the sharpness in temporal, spectral, contextual changes). These observations support the notion of a common neural substrate of change detection for different deviance-related responses.

Chapter 7

Summary, discussion, and future directions

"Trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers: It just cannot be done. In order to understand bird flight, we have to understand aerodynamics; only then do the structure of feathers and the different shapes of birds' wings make sense."

<div align="right">

– David Marr, 1982

</div>

7.1 Summary and discussion

I proposed a *generic deviance detection principle* based on the observation that many deviance-related cortical responses occur without clear evidence of functionally-specific wiring patterns. The principle utilizes the functional term *deviance detection*, which includes *change detection* and *regularity formation*, to decipher the neural dynamics elicited by an abrupt change in the stimuli. I have formulated the characteristics of these functions in Chapters 4,5, and 6. The proposed principle suggests that reciprocal wiring in the cortex gives rise to the emergence of change detectors that respond to abrupt changes in the representation of regularity. With this notion, the deviance-related responses observed in the cortex such as cortical On/Off responses, the cortical OSR, and the MMN can be regarded as change responses at different levels of abstraction. Below, I summarize the simulation works and conclusions regarding these functions.

Change detection (Chapter 4) is achieved by neural populations that are sensitive to the onset/offset of the neighboring neural activity. Change detection is a local and distributed mechanism through reciprocal wiring, with the *transient disinhibition* supporting onset detection, and the *release from inhibition* supporting offset detection (Figures 4.5 and 4.6). I demonstrated that the reciprocal wiring could account for the properties of cortical On/Off responses, which include the diverse temporal profiles across neurons (Figure 4.3) and the distinct onset/offset receptive fields in a single neuron (Figure 4.4). Since inhibition plays an important role, the NMDA-r antagonists reduce the emergence of change detectors by reducing the inhibitory activities. This is because the E-to-I connections are more seriously affected by the NMDA-r antagonists than the E-to-E connections are (Figure 4.7). Besides, synaptic adaptation facilitates not only onset detection but also offset detection (Figure 4.8).

Regularity formation (Chapter 5) is a process that collects and integrates repetitive features over time. Regularity formation relies on the converging connections to some inhibitory populations with a larger time constant τ_i. The neural representation of regularity is relatively stable compared with the change responses. Both regularities of periodicity (i.e., the 'when') and sequence (i.e., the 'what') are *encoded* in the connection patterns through short-term plasticity and *represented* by the inhibitory activities. NMDA-r antagonists affect regularity formation by either reducing the short-term plasticity or breaking the E/I balance.

Deviance detection (Chapter 6) detects changes in the regularity. The deviance detection process (including *regularity formation* and *change detection*) turns the time-varying sensory inputs into hierarchical spatiotemporal features, which enrich the representations of percepts. Compared with the adaptation hypothesis, the generic deviance detection principle emphasizes the role of change detection that is rather an active process than a passive adaptation mechanism. Compared with the prediction hypothesis, the generic deviance detection principle does not rely on a strict hierarchical structure where the roles of top-down and bottom-up signals need to be distinguished. Ideally, a random recurrent network should be flexible and complex enough to foster regularity formation and change detection. Taken together, I claim that the generic deviance detection principle provides a more comprehensive viewpoint than the *adaptation* and the *prediction* hypotheses do. Moreover, the principle can be implemented with biologically meaningful models (e.g., neural mass models), so the model predictions can be verified in experiments.

Although I aimed to reconcile the confusion and conflict between the *adaptation* and the *prediction* hypotheses, the proposed principle provides some opinions and model predictions that are not in agreement with the two hypotheses. For example, the omitted-stimulus responses should not be considered as pure endogenous activities such as a sustained resonance or a prediction signal. Also, the NMDA-r antagonists should generally dampen all deviance-related responses. Below, in Section 7.2, I seek *a unifying framework of perception* that integrates the proposed principle. In Section 7.3, I point out future directions which include experimental validation and theoretical extension.

7.2   Toward a unifying framework of perception

The approaches to investigating how the brain perceives the world can be twofold. The **bottom-up approach** is structure driven, starting from the description of a biological structure, and adding computational insights when necessary. The **top-down approach** is hypothesis driven, starting from the computations needed to be implemented, and adding biological details when necessary.

With a bottom-up approach, I focused on non-specialized networks and proposed that *change detection* and *regularity formation* might serve as basic building blocks in perception. However, how the building blocks are organized to achieve cognitive functions is not clear. As David Marr said in [Marr, 1982], the structure of feathers and wings do not make sense without a connection with aerodynamics. Therefore, it is necessary to integrate the knowledge gained from both bottom-up and top-down approaches in order to understand perception. The ultimate goal is to reach a level of understanding where a computational model can be built with high cognitive fidelity and high biological fidelity (as illustrated in Figure 3 in [Kriegeskorte and Douglas, 2018]).

Below in Session 7.2.1, I briefly describe the bottom-up approaches and some biological constraints at mesoscopic level. In Session 7.2.2, I reviewed some existing top-down approaches, with a focus on the biological fidelity and the compatibility with the generic deviance detection principle. I seek for a unifying framework of perception that considers *deviance detection functions* (i.e., regularity formation and change detection) as basic building blocks.

7.2.1   Bottom-up approaches

The bottom-up approaches look for basic neural computations and properties (e.g., association and competition, feature extraction and selectivity, regularity formation and change detection, signal priming and gating [Kunze et al., 2019], and working memory [Kunze et al., 2017]) in neural networks with more strict biological constraints. These computations and properties are the building blocks that presumably serve emergent cognitive functions such as perception, attention control, working memory, language acquisition/comprehension/production, reasoning, decision making, and creativity.

So far there is still a gap between the building blocks and the cognitive functions. One reason can be that most building blocks suggested by the bottom-up approaches were not well integrated into a unifying framework drawn from the top-down approaches. Below I list the biological constraints considered in the deviance detection functions. These biological constraints are to be used to review the top-down approaches. This reflects, to a certain extent, whether a top-down framework precisely allocates the bottom-up building blocks in a biological sense.

1. **Bidirectional connections**
   Reciprocal connections are commonly observed in the cortex. The network structure should also include bidirectional connections to allow *direct interaction* between two nodes. The feedforward neural network, for example, is not biological in this sense.

2. **Soft hierarchy**
   Even though the brain is likely to process sensory inputs in a hierarchical manner, strict structural hierarchy was not found in the cortex. As concluded in [Hilgetag et al., 2000], the hierarchical ordering of brain areas determined by laminar connections may be indeterminate (i.e., not unique). That is, the reciprocal connections between two cortical areas may not be asymmetrical enough to define their relative positions in the hierarchy. Moreover, the cortical areas are not connected in the form of a pure nest (i.e., there are loops). As Jeff Hawkins said in [Hawkins et al., 2011], "Some regions receive input directly from the senses and other regions receive input only after it has passed through several other regions. It is the region-to-region connectivity that defines the hierarchy." Soft hierarchy, which alters according to

input-dependent region-to-region connectivity, is more biological than strict hierarchy.

3. **Inhibition**

   There are many types of inhibitory neurons in the cortex. The functional roles of inhibition and disinhibition have gained more attention. A network that includes inhibitory neurons is more biological.

4. **Synaptic plasticity**

   It is more biological if a theory uses learning rules based on experimentally observed short-term and long-term plasticity.

7.2.2 Top-down approaches

The top-down approaches start with a hypothesis (e.g., the brain might be solving a Bayesian inference problem), derive the algorithmic computations, and then go further down to its physical implementation (see Marr's levels of analysis [Marr and Poggio, 1976]). For simplicity, I summarized the *block diagram* and the *implementation* (if mentioned) of the top-down approaches. This helps to build bridges between the top-down and the bottom-up approaches.

1. **Block diagram (information flow)**

   The block diagram shows the information flow in a network. The information flow includes inter-node *communication* and intra-node *computation*. Take predictive-coding theory for example. The communication between nodes A and B includes prediction signals (from A to B) and error signals (from B to A). The computation define the task of a node. For example, node A needs to collect affable information to compute the prediction on node B, and node B need to compute prediction error.

2. **Implementation of block diagram (neural coding)**

   For communication, the implementation considers which pathways (e.g., a pathway that originates from layer 5 and ends at layer 2/3) and which coding schemes (e.g., rate coding, temporal coding, oscillations, etc.) are to be utilized. For computation, the implementation considers how the explicit representation (e.g., spikes) and the implicit representation (e.g., synaptic strength) interact to achieve desired computations in a node (canonical microcircuits).

I picked three theories: predictive-coding theory [Friston, 2005], Heeger's theory of cortical function [Heeger, 2017], and hierarchical temporal memory [George and Hawkins, 2009]. These theories proposed different viewpoints on how the brain infers the outer environment from the sensory stimuli. The block diagram and the implementation based on the theories also vary.

1. **Predictive-coding theory**

   The predictive-coding theory hypothesizes that the brain adapts its internal model of the world or *generative model* to match external sensory stimuli through *perceptual inference* (i.e., inferring the causes of sensory input) and *perceptual learning* (i.e., learning the relationship between input and cause). There has been several versions of predictive-coding algorithms proposed to update a generative model from sensory data (see review in [Spratling, 2017]).

   Karl Friston suggested in [Friston, 2005], that the processes of perceptual inference and perceptual learning can be viewed as an inverse problem and can be resolved by minimizing the brain's free energy (or prediction error at all levels of the cortical hierarchy). The evoked cortical responses are transient expressions of prediction error during perceptual inference, and the changes in synaptic efficacy result in perceptual learning. Both processes reflect the brain's attempt to minimize the free energy. The prediction (top-down signal; inhibitory) carries prior information from the high hierarchical levels, and the prediction error (bottom-up signal; excitatory) updates the generative model. Friston's predictive-coding scheme fits with the biological constraints in terms of bidirectional connections and inhibition. Whether this scheme works in a soft hierarchy condition is not verified yet.

   Michael W. Spratling proposed a different viewpoint in grouping the neural populations [Spratling, 2008a, Spratling, 2008b]: The interaction between top-down prediction and bottom-up prediction error is implemented within a cortical area. Therefore, the intra-area connections are inhibitory (top-down) and excitatory (bottom-up), and the inter-area connections can be purely excitatory. The 'prediction populations' in an area send excitatory outputs both to the 'error populations' at the higher hierarchy and to the 'prediction populations' at the lower hierarchy. Spratling's predictive-coding scheme fits with the biological constraints in terms of bidirectional connections, soft hierarchy, and inhibition.

2. **Heeger's theory of cortical function**

David J. Heeger proposed a unified theory of cortical function for guiding neuroscience and artificial intelligence research [Heeger, 2017]. This was done by adding feedback and recurrent connections to a feedforward neural network. The theory demonstrates how information is processed in the network to perform three types of states in perception: *inference* (i.e., interpreting sensory input with a prior or expectation), *exploration* (e.g., binocular rivalry, motion-induced blindness, the Necker cube, and Rubin's face/vase figure), and *prediction* (e.g., predicting motion velocity for behavioral control). The theory was implemented in a recurrent neural network where the network *state parameters* affect the combination of feedforward drive, feedback drive, and prior drive on a node, which results in different network behaviors that are sensory driven, prior driven, or mixed. The state parameters are hypothesized to correspond to attention, acetylcholine (ACh), noradrenaline, and oscillatory activity.

Heeger's theory has a conceptual difference from the predictive-coding theory. The theory does not hypothesize an inversion of an internal model. Instead, the theory suggests that the prediction (i.e., layer responses) propagates forward, and the prediction error (i.e., mismatch between the layer responses and the feedforward drive from the previous layer) propagates backward. (The backward propagation of prediction error has a similar functional role as the backpropagation at the training stage in artificial neural networks.) As opposite to the predictive-coding theory, Heeger's theory showed that perceptual inference may not necessarily be implemented in a Bayesian inference manner. Heeger's theory was demonstrated in a strict hierarchical structure. Whether the theory allows soft hierarchy is not proved yet. The physical implementation was not proposed yet by the theory.

3. **Hierarchical temporal memory**

Hierarchical temporal memory (HTM) [Hawkins et al., 2011] is a machine learning theory developed to capture the structural and algorithmic properties of the cortex. The concept behind the theory is that sensory data usually carries sequential information (or temporal order) at many levels of abstraction, and the cortex builds a model of the world using a spatio-temporal hierarchy. For applications, the HTM networks are used to learn and recognize time-varying data such as a sequence of visual patterns and spoken lan-

guage. According to the HTM theory, *inference* means recognizing spatial and temporal input pattern as similar to previously learned patterns. *Prediction* means activating other regions into a predictive state (or pre-activated state) that will likely become active by future sensory input.

In the HTM network, every node (called HTM region) comprises a coincidence detector and a mixture of Markov chains. An HTM region learns about the world by finding patterns and then forming sequences of patterns from its input. Every HTM region stores sequences of patterns and provides its representation (belief) to other HTM regions. The communication between regions can be viewed as a Bayesian belief propagation technique, where each HTM region provides the belief about itself (bottom-up, feedforward) and its belief over others (top-down, feedback).

For physical implementation, Dileep George and Jeff Hawkins mapped the mathematics of HTM to the microcircuits of cortical columns and provided several predictions [George and Hawkins, 2009]: The neurons in each cortical layer of the microcircuits take care of different functional roles such as detecting coincidence (layer 4), forming sequence memory, incorporating feedback information and sending feedforward information (Layer 2/3), calculating belief distribution over coincidence patterns (layer 5), computing the feedback information for children (layer 6), and broadcasting feedback information (Layer 1).

As for the capability of deviance detection, each HTM region is supposed to be a novelty detector because it "knows" when the input sequence is not expected. However, the theory did not mention how a HTM region responds to a deviant. Moreover, the proposed microcircuit of inference does not contain inhibitory neurons, because the authors suggested that inhibitory neurons are involved only in the learning phase. Based on the structure of the derived building block, I assume the HTM theory can potentially work in a soft hierarchy condition.

The three theories utilize different contents in the information flow (defined above). In predictive-coding theory and Heeger's theory, the information contains prediction and prediction error (in the format of either values or belief). In HTM theory, the information contains self- and other-estimated beliefs over the Markov chains.

The generic deviance detection principle proposed in this work suggests the information contains steady features (regularity) and transient features (change response). The two types of feature can be analogous to prediction and prediction error in predictive-coding theory and Heeger's theory. The steady feature can be as well analogous to the belief in HTM theory, but the counterpart of the transient feature is missing. In this sense, the predictive-coding theory and Heeger's theory both provide a more compatible framework for the deviance detection functions than the HTM theory does. In addition, the proposed principle suggests that the perceptual hierarchy does not arise from a strict hierarchical network structure. A setting of soft hierarchy (described in Section 7.2.1) is also an important criterion for an ideal unifying framework. In this sense, Spratling's predictive-coding theory [Spratling, 2008a, Spratling, 2008b] and the HTM theory [Hawkins et al., 2011] are more biological than Frinston's predictive-coding theory and Heeger's theory. Based on the above comparison, Spratling's predictive-coding theory seems to provide a potential compatible framework to integrate the deviance detection functions.

7.3 Future directions

The future directions include (i) experimental validation, where the model predictions can be validated, and (ii) theoretical extension to approach a unifying theory of perception based on the proposed generic deviance detection principle. I have itemized the two parts as below:

Experimental validation

1. **Similar laminar profiles in cortical deviance-related responses. (model prediction)**
   The cortical Off response, cortical OSR, and MMN should show similar laminar profiles, for example, sink in layer 2/3 [Javitt et al., 1996]. In addition, inhibited activity of inhibitory interneurons near the location of the deviance response should be observed during the stimulus. Taking the pitch MMN as an example (assuming cortical area A has the best frequency of standard tone A, area B has the best frequency of deviant tone B, and area X is the location of MMN), the inhibitory interneurons in area X should be inhibited by tone A. In addition, area X can be a broader area (which may still include area B) that surrounds area A.

2. **Sensitivity to NMDA-r antagonists. (model prediction)**

   It is known that the amplitude of MMN is reduced by NMDA-r antagonists. NMDA-r antagonists can affect deviance responses at the stage of *change detection* by altering the E/I balance (Section 4.3.4). Therefore, the cortical Off response and the cortical OSR should be also sensitive to the NMDA-r antagonists, if the assumption of a common change detection mechanism is correct. Moreover, NMDA-r antagonists can affect deviance responses at the stage of *regularity formation* by reducing the short-term plasticity (Section 5.4). Therefore, the amplitude of entrainment to periodic stimuli in omission paradigms should be reduced by NMDA-r antagonists. Note: this prediction may have been partially supported by impaired delta entrainment in patients with schizophrenia [Lee et al., 2017].

3. **Identification of regularity formation and change detection.**

   In the simulations, I found that larger τ_i is beneficial for regularity formation, while small τ_i is enough for change detection (Section 5.2.2). Therefore, the neural populations serving these two functions should show different temporal profiles that reflect the time constant τ_i (e.g., the oscillating frequencies), which makes it possible to label these neural populations with the corresponding functional roles. The geometric distribution (as well as the laminar profiles) of the two functions may provide evidence to support the generic deviance detection principle, and also provide direct insight into the information flow in the perceptual hierarchy.

Theoretical extensions

1. **To understand the attention mechanism in perception.**

   Attention affects the amplitudes of deviance responses. In this thesis, I have not yet considered the attention mechanism in the generic deviance detection principle. The attention mechanism in perception is still unclear and is often confused with the concept of prediction. Psychological experiments suggested that attention and prediction belong to two different processes (see review in [Schröger et al., 2015]). In the predictive-coding framework [Feldman and Friston, 2010], both attention and prediction have a top-down effect, where prediction suppresses the afferent signal, and attention enhances the gain of the error signal.

I hypothesize that attention is an ensemble product by distributed brain areas. The regularity formation is affected by the global pattern of neural activities (selective attention), and the change response can effectively alter the global pattern (switch of attention). This viewpoint fits the concept of reservoir computing and has higher flexibility (maybe also higher biological fidelity) compared to the predictive-coding concept that emphasizes the separation between top-down and bottom-up signals (as discussed in Section 7.2.2).

2. **Simulations to examine attention effect on regularity formation.**
   The goal is to build a larger neural network to examine the hypothesis of attention mentioned in the previous point. The neural network is to contain more layers (with feedback connections) or more areas (with inter-area connections) where each layer/area represent the regularity of a certain feature. One can then examine how collective firing patterns (i.e., global patterns) affect regularity formation.

3. **To study the various forms of short-term plasticity.**
   The plasticity of the various synaptic types makes the network dynamics more adaptive but also harder to investigate. Although I suggested the need for other form of short-term plasticity beside synaptic adaptation in regularity formation, I have not yet verified the biological fidelity of the plasticity rules (Chapter 5). Short-term plasticity is affected by selective attention [Jääskeläinen et al., 2011, Jääskeläinen and Ahveninen, 2014]. The goal is to use simulations to elaborate the relationship among short-term plasticity, regularity formation, and attention.

4. **To investigate the role of various types of inhibitory populations.**
   The various types of GABAergic neurons in the cortex have different characteristics (e.g., different distribution in layers, different target locations and target neurons, and different spiking patterns, time constant, and plasticity) and are likely to serve different functional roles [Griffen and Maffei, 2014]. Also, the role of disinhibition has been a focus of the studies on perception and memory [Barron et al., 2017] as well as context-dependent behavior [Kuchibhotla et al., 2017, Pi et al., 2013]. To understand the detailed interaction between different inhibitory neurons during the process of deviance detection, one can consider different types of inhibitory populations in the simulation and compare the simulation results with the calcium-imaging studies where the GABAergic neurons expressing different markers such

as parvalbumin (PV), somatostatin (SST), and ionotropic serotonin receptor 5HT3a (5HT3aR) can be separately identified [Rudy et al., 2011, Tremblay et al., 2016].

5. **Simulations of cross-modal omission responses.**

   The cross-modal omission responses have not yet been modeled. Linking the motor-related suppression of auditory responses [Schneider et al., 2018] and the motor-related omission response in the simulation would gain knowledge on the interaction between motor and perception. The simulations may require the consideration of a network of more than two areas (point 2 in Theoretical Extensions) and the assumption of a disinhibition mechanism (point 4 in Theoretical Extensions).

6. **To investigate the emergence of higher cognitive function.**

   Modeling and simulations have the potential to investigate the emergence of higher cognitive functions such as language acquisition and comprehension. By integrating the basic functional properties such as the *working memory* and *signal flow gating* in a three-population model [Kunze et al., 2017], the *conditional priming* in two interacting three-population model [Kunze et al., 2019], and the *regularity formation* and *change detection* demonstrated in this thesis, future modeling studies can potentially investigate the emergence of higher cognitive functions, where the simulated signals can be verified by MEG/EEG observations.

Bibliography

[Amenedo and Escera, 2000] Amenedo, E. and Escera, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *European Journal of Neuroscience*, 12(7):2570–2574.

[Andreou et al., 2015] Andreou, L.-V., Griffiths, T. D., and Chait, M. (2015). Sensitivity to the temporal structure of rapid sound sequences—an meg study. *Neuroimage*, 110:194–204.

[Baba et al., 2016] Baba, H., Tsukano, H., Hishida, R., Takahashi, K., Horii, A., Takahashi, S., and Shibuki, K. (2016). Auditory cortical field coding long-lasting tonal offsets in mice. *Scientific reports*, 6:34421.

[Barascud et al., 2016] Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., and Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences*, 113(5):E616–E625.

[Barron et al., 2017] Barron, H. C., Vogels, T. P., Behrens, T. E., and Ramaswami, M. (2017). Inhibitory engrams in perception and memory. *Proceedings of the National Academy of Sciences*, 114(26):6666–6674.

[Bartha-Doering et al., 2015] Bartha-Doering, L., Deuster, D., Giordano, V., am Zehnhoff-Dinnesen, A., and Dobel, C. (2015). A systematic review of the mismatch negativity as an index for auditory sensory memory: From basic research to clinical and developmental perspectives. *Psychophysiology*, 52(9):1115–1130.

[Barton et al., 2012] Barton, B., Venezia, J. H., Saberi, K., Hickok, G., and Brewer, A. A. (2012). Orthogonal acoustic dimensions define auditory field

maps in human cortex. *Proceedings of the National Academy of Sciences*, 109(50):20738–20743.

[Baumann et al., 2011] Baumann, S., Griffiths, T. D., Sun, L., Petkov, C. I., Thiele, A., and Rees, A. (2011). Orthogonal representation of sound dimensions in the primate midbrain. *Nature neuroscience*, 14(4):423.

[Baumann et al., 2015] Baumann, S., Joly, O., Rees, A., Petkov, C. I., Sun, L., Thiele, A., and Griffiths, T. D. (2015). The topography of frequency and time representation in primate auditory cortices. *Elife*, 4:e03256.

[Behrend et al., 2002] Behrend, O., Brand, A., Kapfer, C., and Grothe, B. (2002). Auditory response properties in the superior paraolivary nucleus of the gerbil. *Journal of neurophysiology*, 87(6):2915–2928.

[Bendixen et al., 2007] Bendixen, A., Roeber, U., and Schröger, E. (2007). Regularity extraction and application in dynamic auditory stimulus sequences. *Journal of Cognitive Neuroscience*, 19(10):1664–1677.

[Bendixen et al., 2009] Bendixen, A., Schröger, E., and Winkler, I. (2009). I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience*, 29(26):8447–8451.

[Boh et al., 2011] Boh, B., Herholz, S. C., Lappe, C., and Pantev, C. (2011). Processing of complex auditory patterns in musicians and nonmusicians. *PLoS One*, 6(7):e21458.

[Bornkessel-Schlesewsky and Schlesewsky, 2019] Bornkessel-Schlesewsky, I. and Schlesewsky, M. (2019). Towards a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in psychology*, 10:298.

[Brannon et al., 2004] Brannon, E. M., Roussel, L. W., Meck, W. H., and Woldorff, M. (2004). Timing in the baby brain. *Cognitive Brain Research*, 21(2):227–233.

[Brewer and Barton, 2016] Brewer, A. A. and Barton, B. (2016). Maps of the auditory cortex. *Annual review of neuroscience*, 39:385–407.

[Bullock et al., 1990] Bullock, T. H., Hofmann, M. H., Nahm, F. K., New, J. G., and Prechtl, J. C. (1990). Event-related potentials in the retina and optic tectum of fish. *Journal of Neurophysiology*, 64(3):903–914.

[Bullock et al., 1994] Bullock, T. H., Karamürsel, S., Achimowicz, J. Z., Mc-Clune, M. C., and Başar-Eroglu, C. (1994). Dynamic properties of human visual evoked and omitted stimulus potentials. *Electroencephalography and clinical neurophysiology*, 91(1):42–53.

[Bullock et al., 1993] Bullock, T. H., Karamürsel, S., and Hofmann, M. H. (1993). Interval-specific event related potentials to omitted stimuli in the electrosensory pathway in elasmobranchs: an elementary form of expectation. *Journal of Comparative Physiology A*, 172(4):501–510.

[Busse and Woldorff, 2003] Busse, L. and Woldorff, M. G. (2003). The erp omitted stimulus response to "no-stim" events and its implications for fast-rate event-related fmri designs. *Neuroimage*, 18(4):856–864.

[Catterall et al., 2013] Catterall, W. A., Leal, K., and Nanou, E. (2013). Calcium channels and short-term synaptic plasticity. *Journal of Biological Chemistry*, 288(15):10742–10749.

[Chimoto et al., 2002] Chimoto, S., Kitama, T., Qin, L., Sakayori, S., and Sato, Y. (2002). Tonal response patterns of primary auditory cortex neurons in alert cats. *Brain research*, 934(1):34–42.

[Colin et al., 2009] Colin, C., Hoonhorst, I., Markessis, E., Radeau, M., De Tourtchaninoff, M., Foucher, A., Collet, G., and Deltenre, P. (2009). Mismatch negativity (mmn) evoked by sound duration contrasts: an unexpected major effect of deviance direction on amplitudes. *Clinical neurophysiology*, 120(1):51–59.

[Cowan et al., 1993] Cowan, N., Winkler, I., Teder, W., and Näätänen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (erp). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(4):909.

[da Silva, 2004] da Silva, F. L. (2004). Functional localization of brain sources using eeg and/or meg data: volume conductor and source models. *Magnetic resonance imaging*, 22(10):1533–1538.

[Dehmel et al., 2002] Dehmel, S., Kopp-Scheinpflug, C., Dörrscheidt, G. J., and Rübsamen, R. (2002). Electrophysiological characterization of the superior paraolivary nucleus in the mongolian gerbil. *Hearing research*, 172(1-2):18–36.

[Deneux et al., 2016] Deneux, T., Kempf, A., Daret, A., Ponsot, E., and Bathellier, B. (2016). Temporal asymmetries in auditory coding and perception reflect multi-layered nonlinearities. *Nature communications*, 7:12682.

[Feldman and Friston, 2010] Feldman, H. and Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4:215.

[Felix et al., 2011] Felix, R. A., Fridberger, A., Leijon, S., Berrebi, A. S., and Magnusson, A. K. (2011). Sound rhythms are encoded by postinhibitory rebound spiking in the superior paraolivary nucleus. *Journal of Neuroscience*, 31(35):12566–12578.

[Forseth et al., 2018] Forseth, K., Hickok, G., Rollo, P., and Tandon, N. (2018). Prediction in human auditory cortex. *bioRxiv*.

[Friston, 2005] Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, 360(1456):815–836.

[George and Hawkins, 2009] George, D. and Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. *PLoS computational biology*, 5(10):e1000532.

[Griffen and Maffei, 2014] Griffen, T. C. and Maffei, A. (2014). Gabaergic synapses: their plasticity and role in sensory cortex. *Frontiers in cellular neuroscience*, 8:91.

[Grunze et al., 1996] Grunze, H. C., Rainnie, D. G., Hasselmo, M. E., Barkai, E., Hearn, E. F., McCarley, R. W., and Greene, R. W. (1996). Nmda-dependent modulation of ca1 local circuit inhibition. *Journal of Neuroscience*, 16(6):2034–2043.

[Guo and Burkard, 2002] Guo, Y. and Burkard, R. (2002). Onset and offset responses from inferior colliculus and auditory cortex to paired noisebursts: inner hair cell loss. *Hearing research*, 171(1-2):158–166.

[Halgren et al., 1995] Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Liégeois, C., Chauvel, P., and Musolino, A. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. i. superior temporal plane and parietal lobe. *Electroencephalography and clinical neurophysiology*, 94(3):191–220.

[Hawkins et al., 2011] Hawkins, J., Ahmad, S., and Dubinsky, D. (2011). Hierarchical temporal memory including htm cortical learning algorithms. Technical report, Numenta Inc., Palo Alto, California, US.

[He, 2003] He, J. (2003). Corticofugal modulation on both on andoff responses in the nonlemniscal auditory thalamus of the guinea pig. *Journal of Neurophysiology*, 89(1):367–381.

[He et al., 1997] He, J., Hashikawa, T., Ojima, H., and Kinouchi, Y. (1997). Temporal integration and duration tuning in the dorsal zone of cat auditory cortex. *Journal of Neuroscience*, 17(7):2615–2625.

[Heeger, 2017] Heeger, D. J. (2017). Theory of cortical function. *Proceedings of the National Academy of Sciences*, 114(8):1773–1782.

[Herholz et al., 2009] Herholz, S. C., Lappe, C., and Pantev, C. (2009). Looking for a pattern: an meg study on the abstract mismatch negativity in musicians and nonmusicians. *BMC neuroscience*, 10(1):42.

[Hilgetag et al., 2000] Hilgetag, C.-C., O'Neill, M. A., and Young, M. P. (2000). Hierarchical organization of macaque and cat cortical sensory systems explored with a novel network processor. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1393):71–89.

[Homann et al., 2017] Homann, J., Koay, S. A., Glidden, A. M., Tank, D. W., and Berry, M. J. (2017). Predictive coding of novel versus familiar stimuli in the primary visual cortex. *bioRxiv*, page 197608.

[Horváth et al., 2010] Horváth, J., Müller, D., Weise, A., and Schröger, E. (2010). Omission mismatch negativity builds up late. *Neuroreport*, 21(7):537–541.

[Hsiao et al., 2010] Hsiao, F.-J., Cheng, C.-H., Liao, K.-K., and Lin, Y.-Y. (2010). Cortico-cortical phase synchrony in auditory mismatch processing. *Biological Psychology*, 84(2):336–345.

[Hsu et al., 2010] Hsu, W.-Y., Cheng, C.-H., Lin, H.-C., Liao, K.-K., Wu, Z.-A., Ho, L.-T., and Lin, Y.-Y. (2010). Memory-based mismatch response to changes in duration of auditory stimuli: An meg study. *Clinical Neurophysiology*, 121(10):1744–1750.

[Jääskeläinen and Ahveninen, 2014] Jääskeläinen, I. P. and Ahveninen, J. (2014). Auditory-cortex short-term plasticity induced by selective attention. *Neural plasticity*, 2014.

[Jääskeläinen et al., 2011] Jääskeläinen, I. P., Ahveninen, J., Andermann, M. L., Belliveau, J. W., Raij, T., and Sams, M. (2011). Short-term plasticity as a neural mechanism supporting memory and attentional functions. *Brain research*, 1422:66–81.

[Jackson and Bolger, 2014] Jackson, A. F. and Bolger, D. J. (2014). The neurophysiological bases of eeg and eeg measurement: A review for the rest of us. *Psychophysiology*, 51(11):1061–1071.

[Jacobsen and Schröger, 2003] Jacobsen, T. and Schröger, E. (2003). Measuring duration mismatch negativity. *Clinical Neurophysiology*, 114(6):1133–1143.

[Jacobsen et al., 2003] Jacobsen, T., Schröger, E., Horenkamp, T., and Winkler, I. (2003). Mismatch negativity to pitch change: varied stimulus proportions in controlling effects of neural refractoriness on human auditory event-related brain potentials. *Neuroscience letters*, 344(2):79–82.

[Jansen and Rit, 1995] Jansen, B. H. and Rit, V. G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological cybernetics*, 73(4):357–366.

[Jansen et al., 1993] Jansen, B. H., Zouridakis, G., and Brandt, M. E. (1993). A neurophysiologically-based mathematical model of flash visual evoked potentials. *Biological cybernetics*, 68(3):275–283.

[Jaramillo et al., 2000] Jaramillo, M., Paavilainen, P., and Näätänen, R. (2000). Mismatch negativity and behavioural discrimination in humans as a function of the magnitude of change in sound duration. *Neuroscience Letters*, 290(2):101–104.

[Javitt et al., 1996] Javitt, D. C., Steinschneider, M., Schroeder, C. E., and Arezzo, J. C. (1996). Role of cortical n-methyl-d-aspartate receptors in auditory sensory memory and mismatch negativity generation: implications for schizophrenia. *Proceedings of the National Academy of Sciences*, 93(21):11962–11967.

[Javitt and Sweet, 2015] Javitt, D. C. and Sweet, R. A. (2015). Auditory dysfunction in schizophrenia: integrating clinical and basic features. *Nature Reviews Neuroscience*, 16(9):535.

[Joachimsthaler et al., 2014] Joachimsthaler, B., Uhlmann, M., Miller, F., Ehret, G., and Kurt, S. (2014). Quantitative analysis of neuronal response properties in primary and higher-order auditory cortical fields of awake house mice (m us musculus). *European Journal of Neuroscience*, 39(6):904–918.

[Karamürsel and Bullock, 1994] Karamürsel, S. and Bullock, T. H. (1994). Dynamics of event-related potentials to trains of light and dark flashes: responses to missing and extra stimuli in elasmobranch fish. *Electroencephalography and clinical neurophysiology*, 90(6):461–471.

[Karamürsel and Bullock, 2000] Karamürsel, S. and Bullock, T. H. (2000). Human auditory fast and slow omitted stimulus potentials and steady-state responses. *International Journal of Neuroscience*, 100(1-4):1–20.

[Knösche et al., 2002] Knösche, T. R., Lattner, S., Maess, B., Schauer, M., and Friederici, A. D. (2002). Early parallel processing of auditory word and voice information. *NeuroImage*, 17(3):1493–1503.

[Kopp-Scheinpflug et al., 2018] Kopp-Scheinpflug, C., Sinclair, J. L., and Linden, J. F. (2018). When sound stops: offset responses in the auditory system. *Trends in neurosciences*, 41(10):712–728.

[Kopp-Scheinpflug et al., 2011] Kopp-Scheinpflug, C., Tozer, A. J., Robinson, S. W., Tempel, B. L., Hennig, M. H., and Forsythe, I. D. (2011). The sound of silence: ionic mechanisms encoding sound termination. *Neuron*, 71(5):911–925.

[Kriegeskorte and Douglas, 2018] Kriegeskorte, N. and Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature neuroscience*, page 1.

[Kuchenbuch et al., 2013] Kuchenbuch, A., Paraskevopoulos, E., Herholz, S. C., and Pantev, C. (2013). Effects of musical training and event probabilities on encoding of complex tone patterns. *BMC neuroscience*, 14(1):51.

[Kuchibhotla et al., 2017] Kuchibhotla, K. V., Gill, J. V., Lindsay, G. W., Papadoyannis, E. S., Field, R. E., Sten, T. A. H., Miller, K. D., and Froemke, R. C.

(2017). Parallel processing by cortical inhibition enables context-dependent behavior. *Nature neuroscience*, 20(1):62.

[Kujala et al., 2001] Kujala, T., Kallio, J., Tervaniemi, M., and Näätänen, R. (2001). The mismatch negativity as an index of temporal processing in audition. *Clinical Neurophysiology*, 112(9):1712–1719.

[Kulesza Jr et al., 2003] Kulesza Jr, R. J., Spirou, G. A., and Berrebi, A. S. (2003). Physiological response properties of neurons in the superior paraolivary nucleus of the rat. *Journal of neurophysiology*.

[Kunze et al., 2019] Kunze, T., Haueisen, J., and Knösche, T. R. (2019). Emergence of cognitive priming and structure building from the hierarchical interaction of canonical microcircuit models. *Biological Cybernetics*, 113(3):273–291.

[Kunze et al., 2017] Kunze, T., Peterson, A. D., Haueisen, J., and Knösche, T. R. (2017). A model of individualized canonical microcircuits supporting cognitive operations. *PLoS ONE*, 12:e0188003.

[Langner et al., 2009] Langner, G., Dinse, H. R., and Godde, B. (2009). A map of periodicity orthogonal to frequency representation in the cat auditory cortex. *Frontiers in integrative neuroscience*, 3:27.

[Large et al., 2010] Large, E. W., Almonte, F. V., and Velasco, M. J. (2010). A canonical model for gradient frequency neural networks. *Physica D: Nonlinear Phenomena*, 239(12):905–911.

[Lee et al., 2017] Lee, M., Sehatpour, P., Hoptman, M. J., Lakatos, P., Dias, E. C., Kantrowitz, J. T., Martinez, A. M., and Javitt, D. C. (2017). Neural mechanisms of mismatch negativity dysfunction in schizophrenia. *Molecular psychiatry*, 22(11):1585.

[Lehmann et al., 2016] Lehmann, A., Arias, D. J., and Schönwiesner, M. (2016). Tracing the neural basis of auditory entrainment. *Neuroscience*, 337:306–314.

[Maaten and Hinton, 2008] Maaten, L. v. d. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605.

[Marr, 1982] Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* Henry Holt and Co., Inc., New York, NY, USA.

[Marr and Poggio, 1976] Marr, D. and Poggio, T. (1976). From understanding computation to understanding neural circuitry.

[May and Tiitinen, 2001] May, P. and Tiitinen, H. (2001). Human cortical processing of auditory events over time. *NeuroReport*, 12(3):573–577.

[May and Tiitinen, 2013] May, P. and Tiitinen, H. (2013). Temporal binding of sound emerges out of anatomical structure and synaptic dynamics of auditory cortex. *Frontiers in computational neuroscience*, 7:152.

[May and Tiitinen, 2010] May, P. J. and Tiitinen, H. (2010). Mismatch negativity (mmn), the deviance-elicited auditory deflection, explained. *Psychophysiology*, 47(1):66–122.

[May et al., 2015] May, P. J., Westö, J., and Tiitinen, H. (2015). Computational modelling suggests that temporal integration results from synaptic adaptation in auditory cortex. *European Journal of Neuroscience*, 41(5):615–630.

[Mustovic et al., 2003] Mustovic, H., Scheffler, K., Di Salle, F., Esposito, F., Neuhoff, J. G., Hennig, J., and Seifritz, E. (2003). Temporal integration of sequential auditory events: silent period in sound pattern activates human planum temporale. *Neuroimage*, 20(1):429–434.

[Näätänen and Kähkönen, 2009] Näätänen, R. and Kähkönen, S. (2009). Central auditory dysfunction in schizophrenia as revealed by the mismatch negativity (mmn) and its magnetic equivalent mmnm: a review. *International Journal of Neuropsychopharmacology*, 12(1):125–135.

[Näätänen et al., 1997] Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615):432.

[Näätänen et al., 1989] Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., and Sams, M. (1989). Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neuroscience letters*, 98(2):217–221.

[Näätänen et al., 2007] Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (mmn) in basic research of central auditory processing: a review. *Clinical neurophysiology*, 118(12):2544–2590.

[Näätänen et al., 2004] Näätänen, R., Syssoeva, O., and Takegata, R. (2004). Automatic time perception in the human brain for intervals ranging from milliseconds to seconds. *Psychophysiology*, 41(4):660–663.

[Ng and Penney, 2014] Ng, K. K. and Penney, T. B. (2014). Probing interval timing with scalp-recorded electroencephalography (eeg). In *Neurobiology of Interval Timing*, pages 187–207. Springer.

[Nishihara et al., 2014] Nishihara, M., Inui, K., Morita, T., Kodaira, M., Mochizuki, H., Otsuru, N., Motomura, E., Ushida, T., and Kakigi, R. (2014). Echoic memory: investigation of its temporal resolution by auditory offset cortical responses. *PloS one*, 9(8):e106553.

[Nishihara et al., 2011] Nishihara, M., Inui, K., Motomura, E., Otsuru, N., Ushida, T., and Kakigi, R. (2011). Auditory n1 as a change-related automatic response. *Neuroscience research*, 71(2):145–148.

[Novitski et al., 2007] Novitski, N., Huotilainen, M., Tervaniemi, M., Näätänen, R., and Fellman, V. (2007). Neonatal frequency discrimination in 250–4000-hz range: Electrophysiological evidence. *Clinical Neurophysiology*, 118(2):412–419.

[Novitski et al., 2004] Novitski, N., Tervaniemi, M., Huotilainen, M., and Näätänen, R. (2004). Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Cognitive Brain Research*, 20(1):26–36.

[Paavilainen, 2013] Paavilainen, P. (2013). The mismatch-negativity (mmn) component of the auditory event-related potential to violations of abstract regularities: a review. *International journal of psychophysiology*, 88(2):109–123.

[Parras et al., 2017] Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdés-Baizabal, C., Escera, C., and Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature communications*, 8(1):2148.

[Phillips et al., 2002] Phillips, D. P., Hall, S., and Boehnke, S. (2002). Central auditory onset responses, and temporal asymmetries in auditory perception. *Hearing research*, 167(1-2):192–205.

[Pi et al., 2013] Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J. I., Huang, Z. J., and Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature*, 503(7477):521.

[Pincze et al., 2001] Pincze, Z., Lakatos, P., Rajkai, C., Ulbert, I., and Karmos, G. (2001). Separation of mismatch negativity and the n1 wave in the auditory cortex of the cat: a topographic study. *Clinical Neurophysiology*, 112(5):778–784.

[Prechtl and Bullock, 1994] Prechtl, J. C. and Bullock, T. H. (1994). Event-related potentials to omitted visual stimuli in a reptile. *Electroencephalography and Clinical Neurophysiology*, 91(1):54–66.

[Qin et al., 2007] Qin, L., Chimoto, S., Sakai, M., Wang, J., and Sato, Y. (2007). Comparison between offset and onset responses of primary auditory cortex on-off neurons in awake cats. *Journal of neurophysiology*.

[Raij et al., 1997] Raij, T., McEvoy, L., Mäkelä, J. P., and Hari, R. (1997). Human auditory cortex is activated by omissions of auditory stimuli. *Brain research*, 745(1-2):134–143.

[Ramón et al., 2001] Ramón, F., Hernández, O. H., and Bullock, T. H. (2001). Event-related potentials in an invertebrate: crayfish emit 'omitted stimulus potentials'. *Journal of experimental biology*, 204(24):4291–4300.

[Recanzone, 2000] Recanzone, G. H. (2000). Response profiles of auditory cortical neurons to tones and noise in behaving macaque monkeys. *Hearing research*, 150(1-2):104–118.

[Regehr, 2012] Regehr, W. G. (2012). Short-term presynaptic plasticity. *Cold Spring Harbor perspectives in biology*, 4(7):a005702.

[Rinne et al., 2006] Rinne, T., Särkkä, A., Degerman, A., Schröger, E., and Alho, K. (2006). Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain research*, 1077(1):135–143.

[Rudy et al., 2011] Rudy, B., Fishell, G., Lee, S., and Hjerling-Leffler, J. (2011). Three groups of interneurons account for nearly 100% of neocortical gabaergic neurons. *Developmental neurobiology*, 71(1):45–61.

[Rujescu et al., 2006] Rujescu, D., Bender, A., Keck, M., Hartmann, A. M., Ohl, F., Raeder, H., Giegling, I., Genius, J., McCarley, R. W., Möller, H.-J., et al.

(2006). A pharmacological model for psychosis based on n-methyl-d-aspartate receptor hypofunction: molecular, cellular, functional and behavioral abnormalities. *Biological psychiatry*, 59(8):721–729.

[Ruusuvirta et al., 2013] Ruusuvirta, T., Lipponen, A., Pellinen, E., Penttonen, M., and Astikainen, P. (2013). Auditory cortical and hippocampal-system mismatch responses to duration deviants in urethane-anesthetized rats. *PloS one*, 8(1):e54624.

[Saha et al., 2017] Saha, D., Sun, W., Li, C., Nizampatnam, S., Padovano, W., Chen, Z., Chen, A., Altan, E., Lo, R., Barbour, D. L., et al. (2017). Engaging and disengaging recurrent inhibition coincides with sensing and unsensing of a sensory stimulus. *Nature communications*, 8:15413.

[Sams et al., 1993] Sams, M., Hari, R., Rif, J., and Knuutila, J. (1993). The human auditory sensory memory trace persists about 10 sec: neuromagnetic evidence. *Journal of cognitive neuroscience*, 5(3):363–370.

[Sams et al., 1985] Sams, M., Paavilainen, P., Alho, K., and Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 62(6):437–448.

[SanMiguel et al., 2013] SanMiguel, I., Saupe, K., and Schröger, E. (2013). I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted" what" but not" when". *Frontiers in human neuroscience*, 7:407.

[Schneider et al., 2018] Schneider, D. M., Sundararajan, J., and Mooney, R. (2018). A cortical filter that learns to suppress the acoustic consequences of movement. *Nature*, 561(7723):391.

[Scholl et al., 2010] Scholl, B., Gao, X., and Wehr, M. (2010). Nonoverlapping sets of synapses drive on responses and off responses in auditory cortex. *Neuron*, 65(3):412–421.

[Schönwiesner et al., 2007] Schönwiesner, M., Novitski, N., Pakarinen, S., Carlson, S., Tervaniemi, M., and Naatanen, R. (2007). Heschl's gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. *Journal of Neurophysiology*.

[Schröger et al., 1994] Schröger, E., Paavilainen, P., and Näätänen, R. (1994). Mismatch negativity to changes in a continuous tone with regularly varying frequencies. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 92(2):140–147.

[Schröger et al., 2015] Schröger, E., Kotz, S. A., and SanMiguel, I. (2015). Bridging prediction and attention in current research on perception and action. *Brain Research*, 1626:1 – 13. Predictive and Attentive Processing in Perception and Action.

[Schwartz et al., 2007] Schwartz, G., Harris, R., Shrom, D., and Berry II, M. J. (2007). Detection and prediction of periodic patterns by the retina. *Nature neuroscience*, 10(5):552.

[Schwartz and Berry II, 2008] Schwartz, G. W. and Berry II, M. J. (2008). Sophisticated temporal pattern recognition in retinal ganglion cells. *Journal of neurophysiology*.

[Shiga et al., 2015] Shiga, T., Althen, H., Cornella, M., Zarnowiec, K., Yabe, H., and Escera, C. (2015). Deviance-related responses along the auditory hierarchy: Combined ffr, mlr and mmn evidence. *PLoS one*, 10(9):e0136794.

[Shiramatsu et al., 2013] Shiramatsu, T. I., Kanzaki, R., and Takahashi, H. (2013). Cortical mapping of mismatch negativity with deviance detection property in rat. *PLoS One*, 8(12):e82663.

[Spaak et al., 2014] Spaak, E., de Lange, F. P., and Jensen, O. (2014). Local entrainment of alpha oscillations by visual stimuli causes cyclic modulation of perception. *Journal of Neuroscience*, 34(10):3536–3544.

[Spiegler et al., 2010] Spiegler, A., Kiebel, S. J., Atay, F. M., and Knösche, T. R. (2010). Bifurcation analysis of neural mass models: Impact of extrinsic inputs and dendritic time constants. *NeuroImage*, 52(3):1041–1058.

[Spiegler et al., 2011] Spiegler, A., Knösche, T. R., Schwab, K., Haueisen, J., and Atay, F. M. (2011). Modeling brain resonance phenomena using a neural mass model. *PLoS computational biology*, 7(12):e1002298.

[Spratling, 2008a] Spratling, M. W. (2008a). Predictive coding as a model of biased competition in visual attention. *Vision research*, 48(12):1391–1408.

[Spratling, 2008b] Spratling, M. W. (2008b). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in computational neuroscience*, 2:4.

[Spratling, 2017] Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and cognition*, 112:92–97.

[Stekelenburg and Vroomen, 2015] Stekelenburg, J. J. and Vroomen, J. (2015). Predictive coding of visual–auditory and motor-auditory events: An electrophysiological study. *Brain research*, 1626:88–96.

[Sussman et al., 2007] Sussman, E. S., Horváth, J., Winkler, I., and Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception & psychophysics*, 69(1):136–152.

[Takahashi et al., 2004] Takahashi, H., Nakao, M., and Kaga, K. (2004). Cortical mapping of auditory-evoked offset responses in rats. *Neuroreport*, 15(10):1565–1569.

[Tervaniemi et al., 2000] Tervaniemi, M., Schröger, E., Saher, M., and Näätänen, R. (2000). Effects of spectral complexity and sound duration on automatic complex-sound pitch processing in humans–a mismatch negativity study. *Neuroscience Letters*, 290(1):66–70.

[Thivierge and Cisek, 2011] Thivierge, J.-P. and Cisek, P. (2011). Spiking neurons that keep the rhythm. *Journal of computational neuroscience*, 30(3):589–605.

[Tiitinen et al., 1994] Tiitinen, H., May, P., Reinikainen, K., and Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, 372(6501):90.

[Toufan et al., 2016] Toufan, R., Moossavi, A., Aghamolaei, M., and Ashayeri, H. (2016). Topographic comparison of mmn to simple versus pattern regularity violations: The effect of timing. *Neuroscience research*, 112:20–25.

[Tremblay et al., 2016] Tremblay, R., Lee, S., and Rudy, B. (2016). Gabaergic interneurons in the neocortex: from cellular properties to circuits. *Neuron*, 91(2):260–292.

[Tse and Penney, 2006] Tse, C.-Y. and Penney, T. B. (2006). Preattentive timing of empty intervals is from marker offset to onset. *Psychophysiology*, 43(2):172–179.

[Uhlhaas and Singer, 2010] Uhlhaas, P. J. and Singer, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. *Nature reviews neuroscience*, 11(2):100.

[Umbricht et al., 2000] Umbricht, D., Schmid, L., Koller, R., Vollenweider, F. X., Hell, D., and Javitt, D. C. (2000). Ketamine-induced deficits in auditory and visual context-dependent processing in healthy volunteers: implications for models of cognitive deficits in schizophrenia. *Archives of general psychiatry*, 57(12):1139–1147.

[van Wassenhove and Lecoutre, 2015] van Wassenhove, V. and Lecoutre, L. (2015). Duration estimation entails predicting when. *Neuroimage*, 106:272–283.

[Volkov and Galazjuk, 1991] Volkov, I. and Galazjuk, A. (1991). Formation of spike response to sound tones in cat auditory cortex neurons: interaction of excitatory and inhibitory effects. *Neuroscience*, 43(2-3):307–321.

[Wacongne, 2016] Wacongne, C. (2016). A predictive coding account of mmn reduction in schizophrenia. *Biological psychology*, 116:68–74.

[Wacongne et al., 2012] Wacongne, C., Changeux, J.-P., and Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience*, 32(11):3665–3678.

[Werner et al., 2008] Werner, B., Cook, P. B., and Passaglia, C. L. (2008). Complex temporal response patterns with a simple retinal circuit. *Journal of neurophysiology*.

[Wilson and Cowan, 1972] Wilson, H. R. and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12(1):1 – 24.

[Winkler et al., 1996] Winkler, I., Karmos, G., and Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain research*, 742(1-2):239–252.

[Xu et al., 2014] Xu, N., Fu, Z.-Y., and Chen, Q.-C. (2014). The function of offset neurons in auditory information processing. *Translational Neuroscience*, 5(4):275–285.

[Yabe et al., 1998] Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Il-moniemi, R. J., and Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, 35(5):615–619.

[Yago et al., 2001a] Yago, E., Corral, M. J., and Escera, C. (2001a). Activation of brain mechanisms of attention switching as a function of auditory frequency change. *Neuroreport*, 12(18):4093–4097.

[Yago et al., 2001b] Yago, E., Escera, C., Alho, K., and Giard, M.-H. (2001b). Cerebral mechanisms underlying orienting of attention towards auditory frequency changes. *Neuroreport*, 12(11):2583–2587.

[Yamashiro et al., 2011] Yamashiro, K., Inui, K., Otsuru, N., and Kakigi, R. (2011). Change-related responses in the human auditory cortex: An meg study. *Psychophysiology*, 48(1):23–30.

[Yamashiro et al., 2009] Yamashiro, K., Inui, K., Otsuru, N., Kida, T., and Kakigi, R. (2009). Automatic auditory off-response in humans: an meg study. *European Journal of Neuroscience*, 30(1):125–131.

[Yaron et al., 2012] Yaron, A., Hershenhoren, I., and Nelken, I. (2012). Sensitivity to complex statistical regularities in rat auditory cortex. *Neuron*, 76(3):603–615.

[Zucker and Regehr, 2002] Zucker, R. S. and Regehr, W. G. (2002). Short-term synaptic plasticity. *Annual review of physiology*, 64(1):355–405.

List of Figures

List of Tables