CITY UNIVERSITY OF HONG KONG
香港城市大學

# Decoding and Prediction of Sound Sequences in the Neocortex
# 皮層對聲音序列信息的解碼和預測

Submitted to
Department of Neuroscience
神經科學系

in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
哲學博士學位

by

Luo Dan
駱丹

February 2022
二零二二年二月

# Abstract

Sequence coding is essential for sensory processing. Encoding information about the probability of a given stimulus being presented, and about the probability of stimuli forming repeated segments, are two critical aspects to abstract the information from the sequences. It is thought that the mammalian auditory system is highly sensitive to the regularities of the sound sequence and able to use predictive processing in order to automatically detect unexpected (deviant) stimuli and segment sound streams. This thesis investigates the predictive processing of sound sequences in the cortex of rodents and humans based on complex sequence regularities.

The first study tests whether neural activity in the rat auditory cortex is modulated by previous segment experience. We recorded subdural responses using electrocorticography from the auditory cortex of 11 anesthetized rats. Prior to recording, four rats were trained to detect familiar triplets of acoustic stimuli (artificial syllables), three were passively exposed to the triplets, and another four rats had no training experience. While low-frequency neural activity was found to entrain (synchronize) to single stimuli, we did not find evidence for entrainment of neural activity to segments (triplets). However, in trained rats (but not in passively exposed and naïve rats), familiar

triplets could be decoded more accurately than unfamiliar triplets based on distributed neural activity in the auditory cortex. These results suggest that rats process acoustic sequences and that the training experience modulates their cortical activity even under subsequent anesthesia.

In the second study, we tested whether the auditory cortical activity in awake mice shows sensitivity to violations of sequences based on local stimulus probability only or whether it is also sensitive to more complex violations based on stimulus order. We employed an auditory oddball paradigm, while recording wide-field calcium imaging from the auditory cortex of awake mice (N=8). During recording, mice were exposed to a series of standard pairs of artificial vowels and several types of deviant pairs. Prior to the oddball condition, we recorded the frequency response areas by presenting pure syllables at different sound levels, which enabled us to dissociate different auditory fields (A1, AAF, and A2). We found that mice could encode both the local probabilities and the more global stimulus patterns and elicited mismatch signals to the substitution deviants (pairs containing novel elements) and transposition deviants (pairs containing familiar elements in the wrong order), but not to omission deviants (pairs with elements missing). Notably, the A2 area elicited larger MMRs to those deviants than A1, which suggests a hierarchical gradient of prediction error signaling in the auditory system.

In the third study, we disambiguate neural correlates of “what” and “when” predictions by independently manipulating the predictability of temporal onset and acoustic contents at two hierarchical levels (single stimuli and stimulus pairs). Healthy volunteers (N=20) performed a repetition detection task while we recorded their neural activity using electroencephalogram. The results reveal that “what” and “when” predictions interactively modulated stimulus-evoked response amplitude in a hierarchically congruent manner, such that faster “when” predictions modulated the amplitude of mismatch responses to unexpected single stimuli, while slower “when” predictions modulated the amplitude of mismatch responses to unexpected stimulus pairs. We also find that the neural effects of these modulations were shared between the two hierarchical levels of prediction signalling in terms of the spatiotemporal distribution of EEG signals. Furthermore, by analyzing the entrainment of low-frequency neural activity to stimulus stream, we found evidence for a gradual increase of entrainment to slow temporal predictions (regarding the timing of stimulus pairs).

In conclusion, the thesis shows that rodents can encode both local probability and more complex and global patterns, such as perceiving as a chunk if they are highly trained. Besides, both rodents and humans can encode transitions and timing knowledge. Our

results suggested that these potential mechanisms for sequence coding (i.e., transitions and timing, chunking) might be evolutionarily conserved in animals and humans.

**CITY UNIVERSITY OF HONG KONG**
**Qualifying Panel and Examination Panel**

| | |
|---|---|
| Surname | LUO |
| First Name | Dan |
| Degree | Doctor of Philosophy |
| College/Department | Department of Neuroscience |

**The Qualifying Panel of the above student is composed of:**

**Supervisor(s)**

| | |
|---|---|
| Prof. Jan W.H. SCHNUPP | Department of Neuroscience<br>City University of Hong Kong |

**Qualifying Panel Member(s)**

| | |
|---|---|
| Dr. LAU Chun Yue Geoffrey | Department of Neuroscience<br>City University of Hong Kong |
| Dr. XIONG Wenjun | Department of Biomedical Sciences<br>City University of Hong Kong |

**This thesis has been examined and approved by the following examiners:**

| | |
|---|---|
| Prof. Jan W.H. SCHNUPP | Department of Neuroscience<br>City University of Hong Kong |
| Dr. CHAN Ho Man | Department of Electrical Engineering, City University of Hong Kong |
| Prof. MALMIERCA Manuel Sánchez | Department of Cell biology and pathology<br>University of Salamanca |
| Dr. DING Nai | College of Biomedical Engineering and Instrument Sciences, Zhejiang University |

# Declaration

I declare that this thesis entitled, ” Decoding and Prediction of Sound Sequences in The Neocortex” represents my original work and the contents of this thesis has not been submitted to this University and other institutions for a degree or any other qualifications in the form of thesis or other report.

Dan Luo

# Dedication

*I dedicate this work to all my friends,*

*and my big family (Yongzhang Luo (1929-2013),*

*Shikun Luo, Liangmei Wen,*

*Chunhua Luo, Zheng Xu, Shicheng Xu,*

*Hemin Feng).*

# Acknowledgments

I would like to express my deepest appreciation to my wise and humorous main supervisor, Prof. Jan W.H. SCHNUPP, who provided me with the opportunity to study in Hong Kong, and my caring and extremely efficient co-supervisor, Dr. Ryszard AUKSZTULEWICZ, who literally taught me everything by hand. I am highly grateful for their academic guidance, kindness, encouragement, patience, continuous support, and attitude toward research and life, which would profoundly influence me in my future career and life. I cannot thank them enough. I am so lucky to have them to be my supervisors of my PhD and it is my great honor to work with them as a student.

Many thanks to all the former and current lab members for their help, Ambika, Alexa, Ahn, Bruno, Cecilia, Chloe, Drew, Fei, Hijee, Jeen, Lucas, Qinglin, Reuben, Shiyi, Tony, and Vani. Wish you all happiness and success.

I am indebted to Prof. Patrick KANOLD at the University of Maryland (now moved to John Hopkins) to supervise and guide the calcium imaging projects and all lab members in his group for their support. Particularly, I sincerely appreciate the senior collaborator in the Kanold lab, Ji LIU, for his supervision and help.

Besides, I am deeply grateful to my committee members, Prof. CHAN Ho Man (City University of Hong Kong), Prof. Manuel S. MALMIERCA (University of Salamanca), and Prof. Nai DING (Zhejiang University), for their valuable time in reading my thesis and giving me constructive comments.

In addition, I would like to extend my deep thanks to Prof. Yi DU (Institute of Psychology, CAS), who literally led me to the auditory research field, for her supervision and support. I also appreciate Prof. Mitch SUTTER (University of California, Davis) for his training and guidance in electrophysiology.

My heartfelt thanks to my friends, Cecilia, Lina, Yue, for helping with my rehearsals. Meanwhile, I am grateful for all my friends who listened to me and encouraged me while I met obstacles and felt I am not researcher material.

Special thanks go to my family members, especially my parents and my sister, for their unconditional love and support. I am obliged to my fiancé, Hemin, and our cat, Liulang, for their presence by my side and for sharing the happiness and pains in my life.

Thanks a lot from my deep heart to all people who inspired and helped me.

In the end, I hope we can defeat the COVID-19 very soon. I wish all the people in this world can live healthily and happily, in love and peace, and they can pursue and achieve what they dream about.

Dan Luo

December 2021

# List of Publications

**Papers** (* and †: equal contribution) **:**

**Dan Luo***, Ji Liu*, Ryszard Auksztulewicz , Tony Ka Wing Yip, Patrick O. Kanold † , Jan W. Schnupp †. *Deviant Processing of Complex Sounds in Mouse Auditory Cortex*. 2022. (in submission)

**Dan Luo**, Kongyan Li , HyunJung An , Jan W. Schnupp †, Ryszard Auksztulewicz †. Learning Boosts the Decoding of Sound Sequences in Rat Auditory Cortex.2021. Current Research in Neurobiology. https://doi.org/10.1016/j.crneur.2021.100019.

Drew Cappotto, **Dan Luo**, HiuWai Lai, Fei Peng, Lucia Melloni, Jan Schnupp†, Ryszard Auksztulewicz†. *Congruency Effects of "What" and "When" Predictions in the Auditory System*. 2022. (in submission)

One further manuscript is in preparation.

**Conference Poster:**

**Dan Luo**, Ryzsard Auksztulewicz, Kongyan Li, Jan Schunupp. *Sequence Learning in Rats: Evidence from Behavioral and ECoG Results*. Poster presented at: Neuroscience 2019. Society for Neuroscience; 2019 Oct 19-23; Chicago, IL.

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **A1** | Primary auditory cortex |
| **A2** | Secondary auditory cortex |
| **AAF** | Auditory field |
| **ABR** | Auditory brainstem response |
| **AC** | Auditory cortex |
| **ACC** | Anterior cingulate cortex |
| **AUC** | Area under the curve |
| **CIC** | Central nucleus of inferior colliculus |
| **CNC** | Cochlear nucleus complex |
| **cpGFP** | Cyclically permutated green fluorescent protein |
| **CV** | Consonant-vowel |
| **DCN** | Dorsal cochlear nucleus |
| **DCIC** | Dorsal cortex of inferior colliculus |
| **ECIC** | External laterally cortex of inferior colliculus |
| **ECoG** | Electrocorticography |
| **EEG** | Electroencephalography |
| **ERP** | Event related potentials |
| **FRA** | Frequency response areas |
| **FWHM** | Full-width-at-half-maximum |
| **GLM** | General linear model |
| **IC** | Inferior colliculus |

| | |
|---|---|
| **IFG** | Inferior frontal gyrus |
| **ITPC** | Intertrial phase coherence |
| **LSO** | Lateral superior olive |
| **MGB** | Medial geniculate body |
| **MGD** | Dorsal division of the medial geniculate body |
| **MGV** | Ventral division of the medial geniculate body |
| **MMN** | Mismatch negativity |
| **MMR** | Mismatch response |
| **MNTB** | Medial nucleus of trapezoid body |
| **MSO** | Medial superior olive |
| **PAF** | Posterior auditory field |
| **rIFG** | Right inferior frontal gyrus |
| **RMS** | Root-mean-square |
| **SOA** | Stimulus-onset asynchrony |
| **SOC** | Superior olivary complex |
| **SRAF** | Suprarhinal auditory field |
| **STG** | Superior temporal gyrus |
| **TDT** | Tucker-Davis Technologies |
| **V1** | Primary visual cortex |
| **VAF** | Ventral auditory field |
| **VCN** | Ventral cochlear nucleus |

# Chapter 1. General Introduction

Detecting the unexpected change of patterns from the dynamic environment is vital for sensory processing, as it is essential to survival for humans and animals. For instance, a small animal looking for food needs to pay attention to surprising sounds in the environment, such as unexpected rustling sounds in the tall grass, to avoid predators, while ignoring ongoing background sounds, like the gurgling water sounds of a nearby stream. At the level of neural mechanisms, this deviance detection process can be indexed by the mismatch response (MMR), which is evoked by unexpected events. In a laboratory setting this is often produced by employing a classical oddball paradigm, when repetitive "standard" stimuli are presented in a sequence, but occasionally interrupted by unexpected "oddball" stimuli, which trigger the MMR (Näätänen et al., 1978). The MMR has been reported across species, including in rodents, cats, and non-human primates, and converging evidence suggests that, rather than reflecting mere neural adaptation to expected stimuli, the MMR reflects an active deviance detection process. which may provide information about higher-order processing in the auditory system (Nelken & Ulanovsky, 2007).

Deviance detection is a prominent example of a more general function of the auditory

system - namely to encode the statistical regularities in its auditory inputs and discover the appropriate rules for segmenting sound streams (Dehaene et al., 2015). The sequencing, parsing, or chunking of auditory input streams is an essential aspect of auditory perception for human and non-human listeners alike. A particularly sophisticated example of auditory processing that owes much to this ability to chunk auditory inputs is the perception of human speech, which is characterized by the ability to generate countless and arbitrarily long sequences (i.e. sentences) using limited and shorter sequences (i.e. syllables and words). Noninvasive recordings of neural activity in humans have identified plausible correlates of chunking. For instance, an earlier electroencephalography (EEG) study (Sanders et al., 2002) suggested that a larger N100 amplitude at the word-onset after training is a neural signature of sequence segmentation and chunking.

Both deviant processing and chunking are thought to be hierarchically organized in the auditory system. Studies to date suggest that, in the case of MMR, the earliest station for predictive processing is found in the non-lemniscal inferior colliculus (Parras et al., 2017). Conversely, the auditory cortex is the earliest region in which predictive processing has been observed for both mismatch responses and for more complex operations such as chunking. Downstream cortical regions, including the right inferior

frontal gyrus (rIFG) and anterior cingulate cortex (ACC), were also reported to involve the higher-order predictive processing (Hofmann-Shen et al., 2020).

The organization of the subcortical and cortical auditory systems are similar across many mammals, including the hierarchy of different regions, their topography, and the laminar distribution of six cortical layers. These structural similarities underlie a functional homology of auditory processing and allow us to use animal models, with various techniques (including invasive electrophysiological and neuroimaging methods as well as genetic tools) typically unavailable in humans, in order to obtain more direct evidence for the neural mechanisms subserving auditory perception, including deviance detection and sequence segmentation. In contrast, non-invasive recording techniques typically used in humans (e.g., electroencephalography; EEG) offer whole-brain coverage and access to more complex cognitive functions which are difficult to operationalize in animal models.

Hence, there are three main sections of related background material in this general introduction chapter. The first section describes the auditory pathway from the peripheral auditory system (including the ear and auditory nerves) to the central auditory system (ascending from the cochlear nucleus complex to the auditory cortex).

The second section of this chapter, dedicated to the significance of sequence coding, reviews two relevant mechanisms: (1) transitions and timing, and (2) chunking. The third section introduces the main recording methods used in this thesis. Following these three background sections, I provide an overview of my thesis and briefly describe each chapter focusing on empirical data.

## The Auditory Pathway

Sound, a mechanical wave, which is emitted from the surface of vibrating objects, propagates through the transmission medium such as air or water. In mammals, there are three sequential stages along the pathway for the hearing process: the outer ear (including pinna and ear canal), the middle ear (including eardrum, tympanic cavity, and ossicles), and the inner ear (mainly including the cochlea) (Figure 1).

### The peripheral auditory system

Besides being the first station to receive the sound wave, the pinna of the outer ear may also help localize the sounds (Batteau, 1967). Subsequently, the sound reaches the outer ear canal and pushes against the eardrum (tympanic membrane) of the middle ear. The sound pressure may be selectively gained in the ear canal due to the resonant cavity comprising the outer ear and eardrum (Aibara et al., 2001; Kurokawa & Goode, 1995). Then the pressure on the eardrum triggers the movement of auditory ossicles, which

acts as an impedance matching of sound traveling in the air-filled ear canal to traveling in the fluid-filled cochlea. The decline of sound pressure during the travel from air to the inner ear could be compensated at this stage. Another function of the middle ear can reduce sound conduction and protect the inner ear if the sound is too loud by the reflex contraction of the middle-ear muscles (Borg & Counter, 1989). The movement of the hair cells on the Corti's organ, which is located on the basilar membrane, releases ion channels which cause changes in membrane voltage gradients and allow the conversion of initial mechanical vibration into electrical signals. These, in turn, stimulate synapses which trigger the excitation of auditory (VIII nerve) fibers (Schnupp et al., 2011).

**Figure 1**

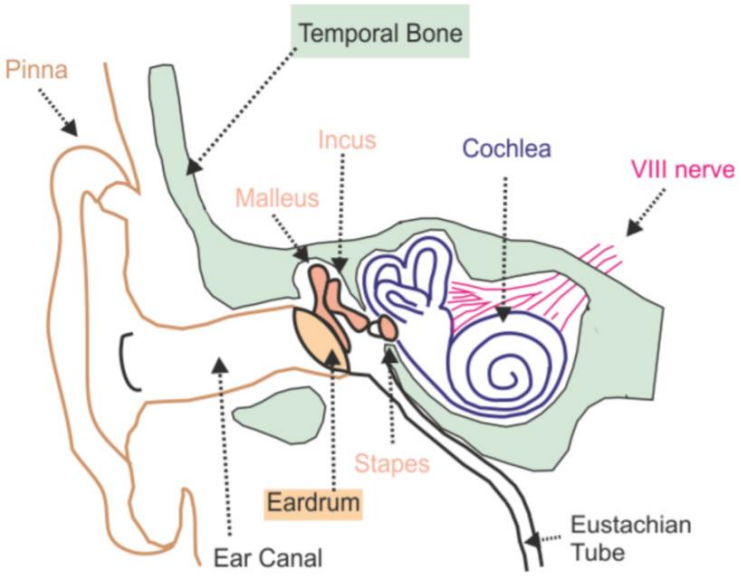


Figure 1. A cross-section of the side of the head, showing structures of the outer, middle, and inner ear. Reprinted from (Schnupp et al., 2011).

## The central auditory system

**Figure 2**

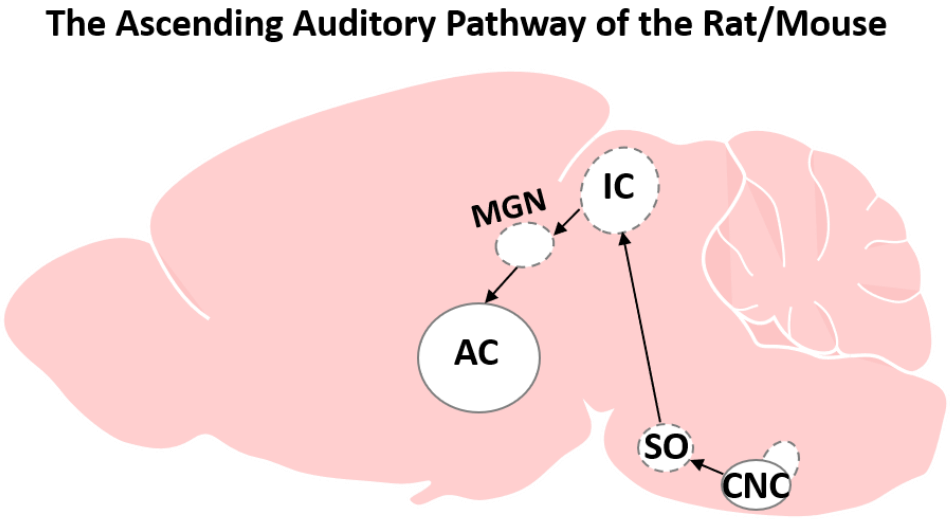


Figure 2. Schematic of ascending auditory pathway of the rat/mouse. CNC: cochlear nucleus complex; SO: superior olivary complex; IC: inferior colliculus; MGN: medial geniculate body; AC: auditory cortex.

### *Cochlear nucleus complex*

The cochlear nucleus complex (CNC) is the first relay center of the ascending central auditory pathway in mammals, consisting of a dorsal cochlear nucleus (DCN) and a ventral cochlear nucleus (VCN). CNC majorly preserves the order of input from auditory nerves according to the topographic organization of frequency encoding (Cant & Benson, 2003; Osen, 1969). Most of the efferent pathways from the CNC terminate in the nuclei of the auditory brainstem. Part of the CNC output directly projects information toward nuclei in the superior olivary complex. Other ascending fibers go

to the lateral lemniscus (Brawer et al., 1974) and the inferior colliculus.

### *Superior Olivary Complex*

The *s*uperior olivary complex (SOC) is the second relay of ascending auditory pathway, which receives afferent projections from the CNC. SOC is a combination of large and small nuclei, and it is the first primary site to encode the binaural differences for spatial hearing (Goldberg & Brown, 1969; Joris et al., 2004). There are three main nuclei in the SOC, including the lateral superior olive (LSO), the medial superior olive (MSO), and the medial nucleus of the trapezoid body (MNTB). The LSO projects bilaterally to the central nucleus of the IC (Shneiderman & Henkel, 1987). The efferent projection from the MSO is mainly toward the inferior colliculus (IC), bypassing the ipsilateral lateral lemniscus (Henkel & Spangler, 1983; Shneiderman & Oliver, 1989). The MNTB is the source of projections to the dorsal portion of the central complex of the lateral lemniscus (Banks & Smith, 1992). The mammalian SOC shows a distinction among species (Malmierca et al., 2003) compared with mammalian CNC. compared with the mammalian CNC. For instance, both the LSO and MNTB in non-primate animals are larger, while those in humans are relatively small (Moore & Moore, 1971).

### *Inferior Colliculus*

The inferior colliculus (IC) in the midbrain is a mandatory auditory relay that converges all direct and indirect inputs from lower auditory centers (Irvine, 1992; Malmierca et al., 2003; Waitzman & Oliver, 2002). There are three major subdivisions in the rat IC (Faye-Lund & Osen, 1985), including the central nucleus (CIC), the dorsal cortex of the IC (DCIC), and (laterally) the external cortex of the IC (ECIC). The CIC mostly receives inputs from the ascending auditory pathway and keeps the tonotopic representation of the sound. The DCIC tends to be most strongly influenced by the top-down projections from the cerebral cortex (Winer et al., 1995). The ECIC receives projections from the cerebral cortex, which is similar to the DCIC. However, the ECIC also receives projections from other non-auditory structures (Coleman & Clerici, 1987; Gruters & Groh, 2012; Yasui et al., 1990). As it is the beginning station of the auditory non-lemniscal pathway, the IC is thought to be the earliest station of predictive processing (Parras et al., 2017).

***The Medial Geniculate Body (auditory thalamus)***

The medial geniculate body (MGB), auditory thalamus, is the 'gateway' to the cortex in the ascending auditory pathway (Hubel & Wiesel, 1961; Sherman & Guillery, 2002). It can be divided into three subdivisions anatomically. The ventral division of the MGB (MGV) is the "core" area for auditory information transmission, and the MGV neurons

can be sharply tuned to frequency response. The dorsal division of the MGB (MGD) is thought to be an integrative area, and the auditory responses in this area can be affected by other non-auditory participations. The MGV and MGD mainly project to the auditory cortex (de la Mothe et al., 2006), while the MGM projects to the auditory cortex and amygdala (Doron & Ledoux, 2000).

***Auditory cortex***

The auditory cortex (AC) is the last relay of the central auditory pathway. All auditory inputs from lower auditory centers converge in the AC which subserves more complex auditory and perceptual processing. The AC can be divided into primary-like areas and secondary areas in most mammals. In rats, there are five main distinct areas of the auditory cortex, including the primary auditory cortex (A1), the anterior auditory field (AAF, also usually considered a "primary" field given its strong, direct thalamic input), and three secondary areas: the ventral auditory field (VAF), the posterior auditory field (PAF), and the suprarenal auditory field (SRAF) (Polley et al., 2007). The most significant distinction of the human and primate cortex is that it is folded (Figure 3) compared with rodents. In the human temporal lobe, the primary auditory cortex (A1) is located along Heschl's gyrus. A1 is tonotopically organized, such that neurons in the anterior area of A1 respond best to low frequencies while the posterior area responds

best to higher frequencies. The surrounding belt area of A1 is the secondary auditory cortex (A2), including the superior temporal gyrus (STG).

Anatomists distinguish two distinct pathways projecting from IC to AC of rodents: the lemniscal auditory pathway and the non-lemniscal pathway. The lemniscal auditory pathway, carrying tonotopically organized and auditory specific information, includes the tonotopically-organized subcortical regions CIC, MGV and the tonotopically-organized areas of the auditory cortex(A1, AAF and VAF). In contrast, the non-lemniscal pathway, the non-tonotopically pathway, includes the DIC, MGD, MGM and the non-tonotopic organized auditory cortex (PAF and SRAF), which plays an important role in sensory integration, temporal pattern recognition, and certain forms of learning (Hu, 2003). Recent reviews (Carbajal & Malmierca, 2017, 2018) suggested that the non-lemniscal pathway constitutes a secondary system capable of encoding more complex auditory information and tracking the past auditory events, and accounts for generating deviance-detection activity and prediction error signals.

Differences in sensitivity of different cortical regions to different components of predictive processing have been shown across species. For instance, a human EEG study (Hofmann-Shen et al., 2020) suggested that the primary auditory cortex and left

inferior frontal gyrus (IFG) are accounting for the “first-order” predictive processing (MMR), which reflect the processing of simple deviant features (such as frequency), while the rIFG and ACC was associated with the “higher-order” processing of novelty, such as the processing of unexpected auditory inputs and regularity violation of known patterns. In rats, the PAF (Parras et al., 2021) is more engaged in context-dependent predictive processing underlying deviance-detection than the other AC fields, suggesting that PAF may function as the chief generator of prediction error signals in the auditory cortex of rats.

**Figure 3**

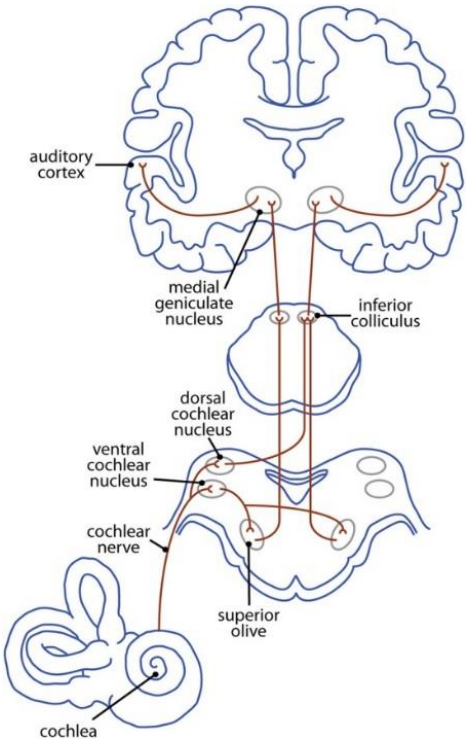


Figure 3. Schematic of the ascending auditory pathway of human, from cochlea to cortex. Figure adapted from (Butler & Lomber, 2013).

## The significance of sequence coding

The mammalian auditory system can encode various types of regularities in the auditory sequence. For sequence coding, (Dehaene et al., 2015) proposed five distinct levels of sequence representation, possibly mediated by different neural mechanisms: (1) transitions and timing knowledge, or the extraction of information about the transitions from one stimulus to the subsequent stimulus; (2) chunking, that is,the grouping of several elements into a single unit; (3) ordinal knowledge, meaning the extraction of information about the sequential order of the stimuli; (4) algebraic patterns, or a formation of abstract schemas that capture the sequential regularities underlying a sequence of stimuli; and (5) nested tree structures, or a representation of suprasegmental (e.g., syntactic) structures characteristic of human languages. To facilitate a comparative perspective, this thesis mainly focuses on the first two mechanisms: transitions and timing knowledge, and chunking (Figure 4), both of which have been previously demonstrated in animal models. Furthermore, to limit the scope of cross-species comparisons, in the thesis I will mainly discuss the neural mechanisms of these functions in rodents and humans.

**Figure 4**

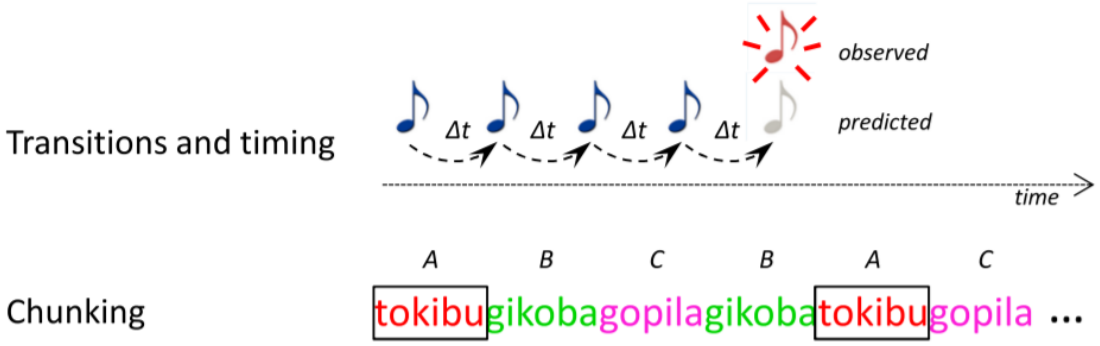


Figure 4. Illustration of the Proposed Taxonomy of Sequence Knowledge. Adapted from (Dehaene et al., 2015).

***Transitions and timing***

An influential integrative theory of perceptual inference, the predictive coding framework (Auksztulewicz & Friston, 2015; Friston, 2005; Friston & Kiebel, 2009), posits that when the brain processes the information about acoustic stimuli presented in a sequence, neurons in the auditory processing pathway are constantly generating and updating the predictions of an internal model of the environment. From this perspective, the primary role of auditory processing is to accurately predict the incoming input and to minimize the discrepancy between the sensory input and the predictive model, conserving energy. However, since the environment is inherently noisy, such discrepancies inevitably arise resulting in prediction errors. These prediction error signals in turn contribute to model updates. Importantly for this thesis, converging

evidence suggests that MMR activity reflects such prediction error signalling (Auksztulewicz & Friston, 2016; May & Tiitinen, 2010).

MMR is modulated not only by the objective probability (high or low) of an entire stimulus, but also by an unexpected change of single acoustic features of the otherwise predicted stimuli. For example, a previous study (Giard et al., 1995) found that the auditory cortex and the prefrontal cortex in humans elicit MMRs to violations of pitch, spectral characteristics, intensity, or duration of the expected stimulus (pure tones). Similar results were also found in animals. In a recent animal study (An et al., 2020), MMR to violations of different acoustic features, including pitch, duration, location, and fine spectral features of the auditory stimuli, were observed using electrocorticographic (ECoG) recordings from the auditory cortex of rats.

Predictions can be formed not only about acoustic features of stimuli, but also about the timing of these stimuli in a sequence. A recent human ECoG study (Auksztulewicz et al., 2018) found that cortical responses to target stimuli were modulated both by the stimulus content expectation, based on past stimulation (75% or 50% transition probability between the previous and current stimuli), and by the onset expectation

based on sequence timing (rhythmic/isochronous or random/jittered). Content and timing predictability modulated activity in dissociable regions and time windows. The "what" predictability (content) increased ECoG response amplitudes over IFG and middle frontal gyri at late latencies (420 – 460 ms), and the "when" predictions selectively enhanced the amplitude of the ECoG response over the precentral gyrus, the supramarginal gyrus, and the rostral middle frontal gyrus both at early (180 ms) and late (430 –450 ms) latencies. Besides, an interaction between "what" and "when" predictability over the posterior STG was observed at an early window. Computational modeling of ECoG data showed that the predictability of content and timing was best explained by dissociable gain control mechanisms in the higher-order auditory regions (STG). Thus, these results suggested that what (content) and when (time) predictions engage complementary neural mechanisms in different cortical regions.

Animal studies also suggest that regularity-based predictive processing is hierarchically organized in the auditory system. (Parras et al., 2017) recorded single-unit activity and local field potentials from anesthetized mice, and found that prediction error signals were present in the major stations of the non-lemniscal auditory pathway of mice and increased along the ascending auditory pathway, which added evidence in support of that the prediction error signals were higher in the non-lemniscal pathway than in the

lemniscal pathway, suggesting that the non-lemniscal pathway works as a higher-order pathway for complex auditory processing (Carbajal & Malmierca, 2018). Using a similar oddball paradigm, in turn, (Lesicko et al., 2021) tested whether inactivation of the auditory cortex can modulate the prediction error in the IC by using optogenetic tools in awake mice and recordings of single neuron activity in the IC. The results show that inactivation of the auditory cortex led to a decrease in prediction error signalling in the IC, which provides more evidence for the top-down effects of predictive processing in the auditory pathway.

Predictive processing is not limited to the auditory modality. A two-photon calcium imaging study (Hamm et al., 2021) identified mismatch responses in calcium activity in the primary visual cortex (V1) of mice, and provided more evidence for hierarchical predictive processing in other modalities. By combining the visual oddball paradigm with calcium imaging in different layers of the primary visual cortex from awake mice, the study found that deviance processing was stronger in supragranular layers (2/3) compared with the deeper layers. Importantly, this study also reported that inactivation of the prefrontal cortex decreased the neural correlates of predictive processing in V1, suggesting that the modulation of predictive

processing in the primary sensory cortex by the higher cortical regions is a modality-general principle.

### *Chunking*

In the mammalian neural system, when a series of contiguous items is frequently presented and repeated in sequence, then these items can be grouped as a coherent segment based on statistical regularities (transition probabilities between stimuli). Studies across species and stimulus domains suggest that neurons in the auditory cortex can encode these groups of items as a 'chunk', which may render the processing of sequences more efficient (Perruchet, 2006). Psycholinguists have proposed that this type of sequence learning ability may play an important part in human language acquisition, especially for linguistic word segmentation. This ability arises quite early in development, as eight-month infants are believed to be able to quickly learn to discriminate novel syllables based on statistical regularities; for instance, infants can detect the difference between novel "part-words" (e.g. tudaro) and "words" (e.g. daropi) after two-hour passive exposure to nonsense syllable streams in which these novel words are embedded (e.g. daropipabikugolatudaropitibudodaropigolatu…) (Saffran et al., 1996). This suggests that chunking might be a relatively automatic process, and a likely precursor of language acquisition (Romberg & Saffran, 2010; Wilson et al., 2017).

More recently, (Ding et al., 2016) found that human subjects' cortical activity is entrained (phase-synchronized) to distinct time scales of speech streams across different linguistic hierarchical levels. Crucially, the entrainment of neural activity to speech streams was modulated by the participants' familiarity with the language in which speech streams are presented. For example, low-frequency neural responses of Chinese listeners exposed to Mandarin speech entrained to the syllable presentation rate as well as to the sentence presentation rate, since they could meaningfully chunk the continuous speech stream into discrete sentences. However, the neural responses of English speakers only entrained to the syllable level but not to the sentence level while listening to Mandarin speech. In a source localization based on ECoG data, results showed that stronger neural entrainment at the sentence level and phrase level were found in the STG, the IFG, and the temporoparietal junction, whereas neural entrainment at the syllable level was stronger in broad cortical areas, including temporal and frontal lobes, as compared with the acoustic control sequences.

A recent ECoG study (Henin et al., 2021) in humans found that neural entrainment arises quickly in both auditory and visual modalities. Neural entrainment to both syllable level and word level were found in STG, motor cortex and pars opercularis of the IFG, while the neural entrainment to words alone was found in IFG and the anterior

temporal lobe. The authors also used multidimensional scaling to infer the types of neural representations based on recordings in the hippocampus, and found that the hippocampal electrodes contained signals related to word identity, whereas the cortical electrodes reflecting the neural entrainment to both words and syllables contained signals related to transitional probability. These results suggested there are multiple parallel computing systems for sequence learning across the cortical-hippocampal circuit. While this result provides a robust neural correlate of the human ability to segment speech streams, it remains an open question whether rodents could use similar chunking mechanisms and/or represent sequences at levels that are hierarchically higher than single elements and their probabilities.

In rodents, an earlier study suggested that rats are sensitive to the statistical co-occurrence of elements in the sequence (Toro & Trobalón, 2005). Artificial grammars were used in this experiment, and two groups of rats were required to learn two grammars respectively. Both grammars consisted of words, part-words and non-words. "Words" were formed as triplets of syllables (e.g. tupiro), and "part-words" were formed by using the first two syllables of a word and the last syllable of another word (e.g. tupidu). "Non-words" were formed by using three syllables which never appeared together in one word. The rats were exposed to a 20-minute syllable stream consisting

only of words before the test sessions. The authors found that the rats could not discriminate part-words from nonwords, but that they could discriminate words from part-words and nonwords, demonstrating that the rat could segment a speech stream by using the overall frequency of the elements (syllables).

The results reported in another study (Murphy et al., 2008) suggested that rodents might have the ability to learn and transfer more abstract rules. In this study, three groups of rats were trained with two visual cues: a bright light (A) and darkness (B), subject to three rules: XYX (e.g., ABA or BAB, presented in a sequential order), XXY, and YXX , respectively. In each group, the rats could get food reinforcement following one stimulus sequence only (e.g. ABA=food, BAB=no food). Results showed that rats in all groups significantly differentiated the reinforced stimulus sequence from the unreinforced stimulus sequence in the last block. As there is no difference between groups, in a second experiment, they trained all rats only on rule 1 (XYX) using three pure tones (ABA, A=3.2 KHz, B=9KHz) and then tested the rats with three novel pure tones with the same rule (low-high-low tone frequency). The rats showed a preference for sounds that were consistent with the rule. These results suggested that rodents can learn and transfer simple rules from stimulus sequences. This study provided direct evidence for sequence learning by rats and suggested rats could transfer their

knowledge from learning experience, like humans. However, until now, there is no consensus regarding the degree to which rodents can learn sequences due to statistical contingencies or explicit rules, and what kinds of neural mechanisms could mediate processing rule-based sequences in rodents.

## Recording techniques

Electrical activities of the neuron or neural population can be recorded through various techniques. For example, when the membrane potential of a neuron is over a certain threshold and elicits an action potential (also called the 'spike'), this high frequency activity can be recorded by the extracellular or intracellular recording. Furthermore, the sub-threshold membrane potential can also be measured by more versatile equipment like patch clamp. However, the brain typically encodes information (i.e., related to complex features of auditory inputs) in the neuronal activity of groups of neurons (Quiroga & Panzeri, 2009). The studies presented in this thesis focus accordingly on recording neuronal populations in different cortical areas by using EEG in humans, as well as electrocorticography (ECoG), and wide-field (WF) calcium imaging in rodents. Figure 5 shows the corresponding temporal and spatial resolution of these techniques.

**Figure 5**

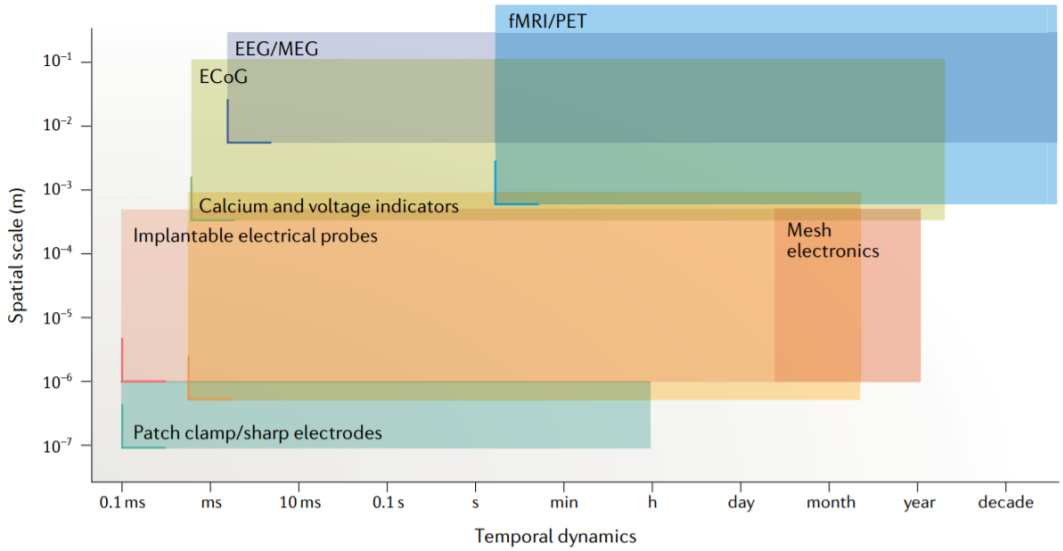


Figure 5. Temporal and spatial resolution of EEG, ECoG and calcium imaging. The horizontal and vertical solid lines denote resolution in temporal and spatial scale, whereas the rectangular color window characterizes the spatiotemporal span of each recording technique. Adapted from (Hong & Lieber, 2019)

### *EEG*

EEG is a widely used method for studying brain electrical activity, especially in human experiments, because of its low cost and non-invasiveness. EEG signals mainly reflect the post-synaptic activity generated by the synchronized activation of large populations of pyramidal neurons in superficial cortical layers (da Silva, 2013). EEG electrodes are usually located on the scalp. However, before the neuronal signals reach the scalp, they must pass through multiple layers of tissues with different electrical properties and complex structures, such as the cerebrospinal fluid, dura mater, skull, and skin. This

means that EEG signals typically have a much lower amplitude and signal-to-noise ratio than intracranial signals (such as those based on ECoG recordings).

In our EEG experiment (Chapter 4), we used an ANT Neuro EEGo Sports 64 channel EEG system which follows the 10-20 system of locating electrodes on the scalp. The electrodes are grounded at the nasion and referenced to the Cpz electrode. EEG signals can be analysed in at least two ways, tapping into complementary kinds of neural activity. In the time domain, EEG signals can be averaged across multiple repeats of the same stimuli to obtain event related potentials (ERP). In humans, ERPs to sensory deviants and standards are commonly contrasted against each other to quantify the so-called mismatch negativity (MMN), a typical example of an MMR linked to signalling prediction errors or rule violations. In the frequency domain, EEG signals can be decomposed into different frequency components to quantify the power and phase consistency of neural oscillations in different frequency bands. Here, we used frequency-domain analyses to calculate the intertrial phase coherence (ITPC) of the EEG signal, in order to reveal the neural entrainment to the spectral peaks of tone sequences at different time scales. Data can also be analysed in a (mass-)univariate manner - i.e., treating each electrode and time point as an independent source of data, performing separate statistical tests for each data point, and accounting for the

correlation between neighbouring electrodes and time points using cluster-based correction for multiple comparisons - as well as in a multivariate manner - i.e., pooling data from multiple electrodes and channels to perform a single statistical test. In particular, multivariate decoding is a useful technique of testing whether the entire pattern of neural activity contains information about e.g. stimulus features. Both univariate and multivariate methods will be used in this thesis.

**Figure 6**

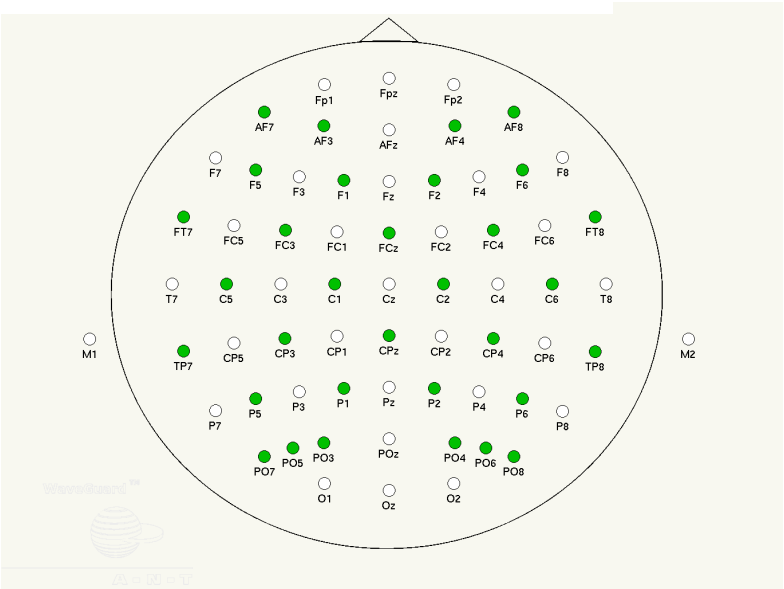


Figure 6. The figure shows the default electrode positions of the international 10-20 system used for EEG recording.

### *ECoG*

ECoG is a popular experimental technique for studying various cortical functions in human clinical research and in animal experiments due to its high spatial and temporal

resolution. ECoG records electrical activity directly on the surface of the cortex, and the signal mainly stems from the neuronal activity in the superficial layers of the cortex. In the cortex, electrical potentials can be measured when spatially aligned pyramidal neurons receive excitatory postsynaptic potentials and then are depolarized in a synchronized way. Similar to EEG signals, the local field potentials in superficial layers constitute the building block of ECoG signals (Buzsáki et al., 2012). Since the neural signals are homologous between EEG and ECoG, data analysis techniques are largely shared between the two data modalities, including time-domain estimation of ERPs and frequency-domain estimation of ITPC, as well as univariate and multivariate analyses.

We employed a flexible 61-channel ECoG array (Insanally et al., 2016) and used it to record electrical activity over the auditory cortex of rats. The ECoG array is arranged on an 8 × 8 square grid (including 3 reference electrodes in the corners of the grid), covering an area of 10.6 mm2. The diameter of the electrode contacts is 203 μm, and the distance between the contacts is 406 μm. Each electrode contact is separately connected with 25 μm wire and spaced at 25 μm, which greatly reduces the width of the entire electrode array. The average electrode impedance is 26.4±1.7 kΩ at 1 kHz, which enables the recording of neural activity at a high signal-noise ratio.

### *WF calcium imaging*

Calcium is an essential intracellular messenger in the mammalian neural system (Grienberger & Konnerth, 2012). Calcium imaging is a widely used technique in neuroscience, based on optical measurements of fluorescence associated with the expression of calcium indicators in the cortex. This allows for measurements of calcium ion concentration, correlated with changes in neuronal activity. There are two different types of calcium indicators: chemical indicators and genetically encoded indicators (GCaMP), and both are commonly applied in rodents, but also in some non-human primate models such as macaques (Li et al., 2017). Calcium imaging is typically characterised by high spatial and temporal resolution.

Calcium ions are an important intracellular messenger in all mammalian cells. Neurons will keep their intracellular Ca++ concentrations very low. Neural activity (action potentials) can lead to the opening of voltage gated Ca++ channels, which triggers rapid changes in intracellular free calcium (Baker et al., 1971; Tank et al., 1988) from a concentration of approx. 100 nM to approx. 10000 nM (Berridge et al., 2000). Most action potentials can be detected at the single neuron level with the newest indicators (GCaMP6s) (Chen et al., 2013) by using two-photon imaging and simultaneous electrophysiology recordings. However, two-photon imaging has a restricted spatial

field of view, limiting the recording to relatively small portions of the cortex. In contrast, wide-field imaging can be used to record from multiple regions simultaneously. While the sensitivity to single action potentials is lower than in two-photon imaging, wide-field imaging nevertheless allows for a recording of approximately 20-30% of all action potentials occurring in a much larger portion of the cortex (Huang et al., 2021). Factors like the sampling rate, resolution of the microscopy, recording duration (causing tissue bleaching), and cranial window size may affect the calcium imaging quality.

Here, we used GCaMP to express green-fluorescent calcium indicators in the brain of mice, specifically in subsets of excitatory neurons. GCaMP is a genetically engineered protein assembled from a cyclically permutated green fluorescent protein subunit, (cpGFP), the calcium-binding protein calmodulin (CaM), and the CaM-interacting M13 peptide [56]. The expression of green-fluorescent calcium indicators (GCaMP6s) is driven by the mouse Thy1 promoter. The transgenic (Thy1-GCaMP6s) mice can be used for calcium imaging to record neuronal activity. GCaMP6s is widely used in research as the expression levels of the indicator in the cortex are quite stable and sufficient for in-vivo imaging (Dana et al., 2014).

## Thesis Overview

This thesis investigates the predictive processing of sound sequences in the cortex of rodents and humans based on complex sequence regularities. In this introductory chapter, I have provided a general introduction of the background, including an overview of the auditory pathways, relevant literature on predictive processing, and recording techniques.

The second chapter covers the first empirical study of my thesis (Ref). In the study, we tested whether neural activity in the rat auditory cortex is modulated by previous experience with stimulus sequences. We trained 4 female Wistar rats to be familiarized with syllable triplets using two alternative forced choice tasks (2AFC) tasks, and recorded ECoG signals from the auditory cortex of 11 animals in total (the 4 trained rats, as well as 4 passively exposed rats and 3 naïve rats). We found that low-frequency neural activity peaks of ECoG in the auditory cortex were observed both at the syllable level and triplet level (consistent with human studies) but did not robustly differentiate between trained and naïve animals. However, in trained rats, but not in passively exposed or naïve rats, could familiar triplets be decoded more accurately than unfamiliar triplets based on the spatial pattern of neural activity in the auditory cortex.

In the third chapter, I examine whether the auditory cortical activity in awake mice shows sensitivity to violations of sequences based on local stimulus probability only or

whether it is also sensitive to more complex violations based on stimulus order. To this end, we recorded wide-field calcium imaging from the auditory cortex of Thy1-GCamp6s transgenic mice exposed to a series of standard pairs of artificial vowels (‘bi pe’) as well as several types of deviant pairs. We found that auditory cortical activity encoded both the local probabilities and the more global stimulus patterns, and elicited mismatch signals to the substitution deviants (‘pe da’ and ‘da bi’) and transposition (‘pe bi’) deviants, but not to omission deviants (‘pe _’). Interestingly, the A2 area elicited more pronounced MMRs to those deviants than A1, which suggests a hierarchical gradient of prediction error signaling.

In Chapter 4, I investigate the neural connection between "what" and "when" predictions by independently manipulating the contents and timing of auditory sequences at two hierarchical levels. Human EEG were recorded while participants were exposed to stimulus sequences containing acoustic oddballs at a hierarchically lower level (unexpected single tones) and a hierarchically higher level (unexpected tone pairs). In an ERP analysis, we quantified the effects of “what” and “when” prediction mechanisms at both the faster and slower time scales. We found an interaction effect between “what” and “when” predictions in terms of their modulation of the MMN amplitude at both time scales, indicating a congruency effect in predictive processing between so-called “what” and “when” events in the auditory system. ITPC analysis

showed that entrainment to stimuli at the slower temporal scale gradually increased over the course of the experiment and was mostly expressed over the right hemisphere, which is consistent with existing research on hemispheric asymmetries between processing at slower and faster time scales (Moser et al., 2021).

Finally, the last chapter summarizes the main findings of my studies.

## References


Aibara, R., Welsh, J. T., Puria, S., & Goode, R. L. (2001). Human middle-ear sound transfer function and cochlear input impedance. *Hearing research, 152*(1-2), 100-109.

Auksztulewicz, R., & Friston, K. (2015). Attentional enhancement of auditory mismatch responses: a DCM/MEG study. *Cerebral cortex, 25*(11), 4273-4283.

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex, 80*, 125-140.

Auksztulewicz, R., Schwiedrzik, C. M., Thesen, T., Doyle, W., Devinsky, O., Nobre, A. C., Schroeder, C. E., Friston, K. J., & Melloni, L. (2018). Not all predictions are equal:"What" and "when" predictions modulate activity in auditory cortex through different mechanisms. *Journal of Neuroscience, 38*(40), 8680-8693.

Baker, P., Hodgkin, A., & Ridgway, E. (1971). Depolarization and calcium entry in squid giant axons. *The Journal of physiology, 218*(3), 709-755.

Banks, M. I., & Smith, P. H. (1992). Intracellular recordings from neurobiotin-labeled cells in brain slices of the rat medial nucleus of the trapezoid body. *Journal of Neuroscience, 12*(7), 2819-2837.

Batteau, D. W. (1967). The role of the pinna in human localization. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 168*(1011), 158-180.

Berridge, M. J., Lipp, P., & Bootman, M. D. (2000). The versatility and universality of calcium signalling. *Nature reviews Molecular cell biology, 1*(1), 11-21.

Borg, E., & Counter, S. A. (1989). The middle-ear muscles. *Scientific American, 261*(2), 74-81.

Brawer, J. R., Morest, D. K., & Kane, E. C. (1974). The neuronal architecture of the cochlear nucleus of the cat. *Journal of Comparative Neurology, 155*(3), 251-299.

Butler, B. E., & Lomber, S. G. (2013). Functional and structural changes throughout the auditory system following congenital and early-onset deafness: implications for hearing restoration. *Frontiers in systems neuroscience, 7*, 92.

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nature reviews neuroscience, 13*(6), 407-420.

Cant, N. B., & Benson, C. G. (2003). Parallel auditory pathways: projection patterns of the different neuronal populations in the dorsal and ventral cochlear nuclei. *Brain research bulletin, 60*(5-6), 457-474.

Carbajal, G. V., & Malmierca, M. S. (2017). The unique role of the non-lemniscal pathway on stimulus-specific adaptation (SSA) in the auditory system. Proceedings of the International Symposium on Auditory and Audiological Research,

Carbajal, G. V., & Malmierca, M. S. (2018). The neuronal basis of predictive coding along the auditory pathway: from the subcortical roots to cortical deviance detection. *Trends in hearing, 22*, 2331216518784822.

Chen, T.-W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., Schreiter, E. R., Kerr, R. A., Orger, M. B., & Jayaraman, V. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature, 499*(7458), 295-300.

Coleman, J. R., & Clerici, W. J. (1987). Sources of projections to subdivisions of the inferior colliculus in the rat. *Journal of Comparative Neurology, 262*(2), 215-226.

da Silva, F. L. (2013). EEG and MEG: relevance to neuroscience. *Neuron, 80*(5), 1112-1128.

Dana, H., Chen, T.-W., Hu, A., Shields, B. C., Guo, C., Looger, L. L., Kim, D. S., & Svoboda, K. (2014). Thy1-GCaMP6 transgenic mice for neuronal population imaging in vivo. *PloS one, 9*(9), e108697.

de la Mothe, L. A., Blumell, S., Kajikawa, Y., & Hackett, T. A. (2006). Cortical

connections of the auditory cortex in marmoset monkeys: core and medial belt regions. *Journal of Comparative Neurology, 496*(1), 27-71.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015, Oct 7). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron, 88*(1), 2-19. https://doi.org/10.1016/j.neuron.2015.09.019

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature neuroscience, 19*(1), 158-164.

Doron, N. N., & Ledoux, J. E. (2000). Cells in the posterior thalamus project to both amygdala and temporal cortex: A quantitative retrograde double-labeling study in the rat. *Journal of Comparative Neurology, 425*(2), 257-274.

Faye-Lund, H., & Osen, K. K. (1985). Anatomy of the inferior colliculus in rat. *Anatomy and embryology, 171*(1), 1-20.

Friston, K. (2005, Apr 29). A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci, 360*(1456), 815-836. https://doi.org/10.1098/rstb.2005.1622

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society B: Biological sciences, 364*(1521), 1211-1221.

Goldberg, J. M., & Brown, P. B. (1969). Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *Journal of neurophysiology, 32*(4), 613-636.

Grienberger, C., & Konnerth, A. (2012). Imaging calcium in neurons. *Neuron, 73*(5), 862-885.

Gruters, K. G., & Groh, J. M. (2012). Sounds and beyond: multisensory and other non-auditory signals in the inferior colliculus. *Frontiers in neural circuits, 6*, 96.

Hamm, J. P., Shymkiv, Y., Han, S., Yang, W., & Yuste, R. (2021). Cortical ensembles selective for context. *Proceedings of the National Academy of Sciences, 118*(14).

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., Doyle, W., Devinsky, O., & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances, 7*(8), eabc4530.

Henkel, C. K., & Spangler, K. M. (1983). Organization of the efferent projections of the medial superior olivary nucleus in the cat as revealed by HRP and autoradiographic tracing methods. *Journal of Comparative Neurology, 221*(4), 416-428.

Hofmann-Shen, C., Vogel, B. O., Kaffes, M., Rudolph, A., Brown, E. C., Tas, C., Brüne, M., & Neuhaus, A. H. (2020). Mapping adaptation, deviance detection, and prediction error in auditory processing. *Neuroimage, 207*, 116432.

Hong, G., & Lieber, C. M. (2019). Novel electrode technologies for neural recordings. *Nature reviews neuroscience, 20*(6), 330-345.

Hu, B. (2003). Functional organization of lemniscal and nonlemniscal auditory thalamus. *Experimental brain research, 153*(4), 543-549.

Huang, L., Ledochowitsch, P., Knoblich, U., Lecoq, J., Murphy, G. J., Reid, R. C., de Vries, S. E., Koch, C., Zeng, H., & Buice, M. A. (2021). Relationship between simultaneously recorded spiking activity and fluorescence signal in GCaMP6 transgenic mice. *Elife, 10*, e51675.

Hubel, D. H., & Wiesel, T. N. (1961). Integrative action in the cat's lateral geniculate body. *The Journal of physiology, 155*(2), 385-398.

Insanally, M., Trumpis, M., Wang, C., Chiang, C.-H., Woods, V., Palopoli-Trojani, K., Bossi, S., Froemke, R. C., & Viventi, J. (2016). A low-cost, multiplexed μECoG system for high-density recordings in freely moving rodents. *Journal of neural engineering, 13*(2), 026030.

Irvine, D. R. (1992). Physiology of the auditory brainstem. In *The mammalian auditory pathway: Neurophysiology* (pp. 153-231). Springer.

Joris, P., Schreiner, C., & Rees, A. (2004). Neural processing of amplitude-modulated

sounds. *Physiological reviews, 84*(2), 541-577.

Kurokawa, H., & Goode, R. L. (1995). Sound pressure gain produced by the human middle ear. *Otolaryngology—Head and Neck Surgery, 113*(4), 349-355.

Lesicko, A. M., Angeloni, C. F., Blackwell, J. M., De Biasi, M., & Geffen, M. N. (2021). Cortico-Fugal Regulation of Predictive Coding. *bioRxiv*.

Li, M., Liu, F., Jiang, H., Lee, T. S., & Tang, S. (2017). Long-term two-photon imaging in awake macaque monkey. *Neuron, 93*(5), 1049-1057. e1043.

Malmierca, M. S., Hernández, O., Falconi, A., Lopez-Poveda, E. A., Merchán, M., & Rees, A. (2003). The commissure of the inferior colliculus shapes frequency response areas in rat: an in vivo study using reversible blockade with microinjection of kynurenic acid. *Experimental brain research, 153*(4), 522-529.

May, P. J., & Tiitinen, H. (2010, Jan 1). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology, 47*(1), 66-122. https://doi.org/10.1111/j.1469-8986.2009.00856.x

Moore, J. K., & Moore, R. (1971). A comparative study of the superior olivary complex in the primate brain. *Folia Primatologica, 16*(1-2), 35-51.

Moser, J., Batterink, L., Li Hegner, Y., Schleger, F., Braun, C., Paller, K.A., Preissl, H., 2021. Dynamics of nonlinguistic statistical learning: From neural entrainment to the emergence of explicit knowledge. Neuroimage 240, 118378. https://doi.org/10.1016/j.neuroimage.2021.118378

Murphy, R. A., Mondragón, E., & Murphy, V. A. (2008). Rule learning by rats. *Science, 319*(5871), 1849-1851.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta psychologica, 42*(4), 313-329.

Nelken, I., & Ulanovsky, N. (2007). Mismatch negativity and stimulus-specific adaptation in animal models. *Journal of Psychophysiology, 21*(3-4), 214-223.

Osen, K. K. (1969). Cytoarchitecture of the cochlear nuclei in the cat. *Journal of*

*Comparative Neurology, 136*(4), 453-483.

Parras, G. G., Casado-Román, L., Schröger, E., & Malmierca, M. S. (2021). The posterior auditory field is the chief generator of prediction error signals in the auditory cortex. *Neuroimage, 242*, 118446.

Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdes-Baizabal, C., Escera, C., & Malmierca, M. S. (2017, Dec 15). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nat Commun, 8*(1), 2148. https://doi.org/10.1038/s41467-017-02038-6

Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdés-Baizabal, C., Escera, C., & Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature communications, 8*(1), 1-17.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. Trends in cognitive sciences, 10(5), 233-238.

Polley, D. B., Read, H. L., Storace, D. A., & Merzenich, M. M. (2007). Multiparametric auditory receptive field organization across five cortical fields in the albino rat. *Journal of neurophysiology, 97*(5), 3621-3638.

Quiroga, R. Q., & Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nature reviews neuroscience, 10*(3), 173-185.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 906-914.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Schnupp, J., Nelken, I., & King, A. (2011). *Auditory neuroscience: Making sense of sound*. MIT press.

Sherman, S. M., & Guillery, R. (2002). The role of the thalamus in the flow of information to the cortex. *Philosophical Transactions of the Royal Society of*

*London. Series B: Biological Sciences, 357*(1428), 1695-1708.

Shneiderman, A., & Henkel, C. K. (1987). Banding of lateral superior olivary nucleus afferents in the inferior colliculus: a possible substrate for sensory integration. *Journal of Comparative Neurology, 266*(4), 519-534.

Shneiderman, A., & Oliver, D. L. (1989). EM autoradiographic study of the projections from the dorsal nucleus of the lateral lemniscus: a possible source of inhibitory inputs to the inferior colliculus. *Journal of Comparative Neurology, 286*(1), 28-47.

Tank, D. W., Sugimori, M., Connor, J. A., & Llinas, R. R. (1988). Spatially resolved calcium dynamics of mammalian Purkinje cells in cerebellar slice. *Science, 242*(4879), 773-777.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & psychophysics, 67*(5), 867-875.

Waitzman, D. M., & Oliver, D. L. (2002). Midbrain.

Wilson, B., Marslen-Wilson, W. D., & Petkov, C. I. (2017). Conserved sequence processing in primate frontal cortex. *Trends in neurosciences, 40*(2), 72-82.

Yasui, Y., Cechetto, D. F., & Saper, C. B. (1990). Evidence for a cholinergic projection from the pedunculopontine tegmental nucleus to the rostral ventrolateral medulla in the rat. *Brain research, 517*(1-2), 19-24.

# Chapter 2. Learning Boosts the Decoding of Sound Sequences in Rat Auditory Cortex


## Abstract

Continuous acoustic streams, such as speech signals, can be chunked into segments containing reoccurring patterns (e.g., words). Noninvasive recordings of neural activity in humans suggest that chunking is underpinned by low-frequency cortical entrainment to the segment presentation rate, and modulated by prior segment experience (e.g., words belonging to a familiar language). Interestingly, previous studies suggest that also primates and rodents may be able to chunk acoustic streams. Here, we test whether neural activity in the rat auditory cortex is modulated by previous segment experience. We recorded subdural responses using electrocorticography (ECoG) from the auditory cortex of 11 anesthetized rats. Prior to recording, four rats were trained to detect familiar triplets of acoustic stimuli (artificial syllables), three were passively exposed to the triplets, while another four rats had no training experience. While low-frequency neural activity peaks were observed at the syllable level, no triplet-rate peaks were observed. Notably, in trained rats (but not in passively exposed and naïve rats), familiar triplets could be decoded more accurately than unfamiliar triplets based on neural activity in the auditory cortex. These results suggest that rats process acoustic sequences, and that

their cortical activity is modulated by the training experience even under subsequent anesthesia.

## Introduction

Chunking of auditory input streams is an essential aspect of auditory perception for human and non-human listeners alike. It is thought that the mammalian auditory system discovers the appropriate rules for segmenting sound streams from observing statistical regularities in its auditory inputs (Dehaene et al., 2015). Developmental studies in humans have shown that infants could discriminate transitional probabilities in auditory sequences after brief passive exposure to statistically regular stimulus streams (Saffran et al., 1996), which suggests that chunking might be a relatively automatic process and a likely precursor of language acquisition (Romberg & Saffran, 2010; Wilson et al., 2017). Noninvasive recordings of neural activity in humans (Ding et al., 2015; 2017) have identified plausible correlates of auditory chunking, relying on the entrainment (phase-synchronization) of cortical activity to distinct time scales of speech streams (monosyllabic words, two-word syntactic phrases, and longer sentences). Crucially, the neural entrainment to speech streams was modulated by the participants' familiarity with the language in which speech was presented. While these results provide a robust neural correlate of the human ability to segment speech streams, it is unknown whether

animals could use similar chunking mechanisms to process familiar stimulus sequences.

In nonhuman primate models, sequence processing has typically been studied using artificial grammar paradigms and focused on cross-species comparisons. In the auditory domain, neural oscillatory signatures of sequence learning in the auditory cortex have been shown to be largely conserved between macaques and humans (Kikuchi et al., 2017), although unlike in the human studies (Batterink & Paller, 2019; Ding et al., 2015), the low-frequency entrainment effects were not frequency-specific to the stimulus presentation rates used in the artificial grammar sequences. Behavioral studies identified differences between species. For instance, Jiang et al. (2018) demonstrated that macaques could learn center-embedded relationships, such as ABC|CBA, in visuospatial sequences after extensive training, while human infants were sensitive to such artificial grammar structures after only brief exposure. In an earlier study (Wilson et al., 2013), which showed that macaques could differentiate more complex auditory artificial grammar structures relative to marmosets, which could detect only simple structure violations, e.g., based on the first position in a sequence.

In rodents, the evidence for learning temporal regularities based on stimulus sequences is mixed. While one earlier study suggested that rats cannot extract abstract rules from

stimulus sequences but are sensitive to the statistical co-occurrence of elements within the sequence (Toro & Trobalón, 2005), another study (Murphy et al., 2008) found that rodents can be trained to learn abstract rules from stimulus sequences and transfer them to new stimuli. Moreover, Bale et al. (2017) found that mice could be trained to discriminate tactile temporal sequences based on stimulus order only. Another study (Gavornik & Bear, 2014) provided electrophysiological evidence of sequence learning in awake mice, namely that the primary visual cortex of trained (but not naïve) animals showed larger evoked responses and higher peak firing rates for familiar vs. unfamiliar visual sequences. These results suggest that cortical neural activity of rodents could be shaped by previous experience of (visual) sequences. However, whether similar effects generalize to the auditory modality, and to what extent animals need to be trained to show sensitivity to sequences, remains unclear.

Here, we recorded neural activity in response to auditory sequences in anesthetized rats (N = 11). Prior to recording, the animals were split into three groups: a trained group was familiarized with syllable sequences using operant conditioning; a passive group received passive exposure to the same sequences; finally, a naïve group did not receive any exposure to these sequences. Using ECoG recordings from the auditory cortex, we quantified neural entrainment to segments of different lengths (syllables and triplets),

tested whether entrained signals can be found in cortical responses of anesthetized rats, and whether they can be modulated by either active or passive auditory experience. Finally, using multivariate decoding of acoustic segments based on spatiotemporal patterns of neural activity, we tested whether stimulus decoding is modulated by prior experience. In this manner, we were able to identify a new neural signature of acoustic stream segmentation in anesthetized rodents.

## Materials and Methods

Eleven female Wistar rats were used in this study. The hearing threshold of all rats was normal, as confirmed by auditory brainstem response (ABR) recordings. Rats in the training group (n = 4) started behavioral training at 8 weeks of age, and once they finished training, they were subject to ECoG recordings. Another, passive group of rats (n = 3) was exposed to prolonged continuous acoustic stimulation prior to ECoG recordings, without any additional training. Finally, a group of naive rats (n = 4) without any training experience were only subject to ECoG recordings. The mean age (± SD) at the time of ECoG recordings was 28.58 ± 7.83 weeks, and the weight 306.36 ± 43.36 g. All experimental procedures were approved by the Animal Research Ethics Sub-Committee of the City University of Hong Kong under the animal license No. (17-76) in DH/SHS/8/2/5 Pt.1.

Artificial acoustic syllables selected from a database of Consonant-Vowel syllables (Ives et al., 2005) were used in behavioral training and subsequent recordings (see Fig. 7A). The speech syllables were analyzed and resynthesized by an open-source vocoder, STRAIGHT (Kawahara, 2006) for MATLAB R2018b (MathWorks), to match the stimulus onset and duration of all syllables (531 ms) and shift the fundamental frequency and formant scalar of each Consonant-Vowel (CV) syllable upward 1 octave to match the optimal rat hearing range (Kelly & Masterton, 1977).

Animals in the trained group performed two-alternative-forced-choice behavior tasks using drinking water as a positive reinforcer. The equipment and task were adapted from (Li et al., 2019). Behavioral training consisted of three separate training stages: a syllable learning stage (training rats to discriminate entire syllables and scrambled syllables), a triplet learning stage (to discriminate entire triplets and scrambled triplets), and a triplet discrimination training stage (to discriminate familiar triplets and novel unfamiliar triplets). In the first two stages (Fig. 7B), we used four syllable triplets as "familiar" stimuli (/pisore/, /gusima/, /dazolu/, /pekina/) and scrambled stimuli as "unfamiliar" stimuli (generated by epoching the waveform of each syllable into 10 segments of equal length, and shuffling the segments; Fig. 7A). In the third and final training stage, we used the same four syllable triplets as "familiar" stimuli and other

syllable triplets (e.g., /gukewo/) as “unfamiliar stimuli”. Once the performance of animals in a particular training stage approached 75% accuracy, they progressed to the next stage. Rats performed approximately 300 trials per session and one session per day. The total duration of training was 33.15 ± 0.52 weeks (mean ± SD). To test whether this ability could transfer to novel stimuli, three animals were tested in a behavioral session using familiar stimuli and previously unheard unfamiliar stimuli.

Animals in the passive group were exposed to continuous presentation of “familiar” triplets (/pisore/, /gusima/, /dazolu/, /pekina/) for 24 hours immediately prior to the electrophysiological recordings. Triplets were presented in a random order through loudspeakers at 80-85 dB in a sound-attenuated box. No further behavior or task was required.

In this manner we created three cohorts of rats. The trained group had experienced the target syllable triplets as reinforced stimuli, the passive exposure group had been exposed to the target triplets for 24 hours, and the naïve group had no experience at all of the target triplets. All three groups were then subjected to identical electrophysiological recordings. There were two conditions in the ECoG experiment, delivered in separate trials. The order of the trials was randomized across rats. In the

first (familiar) condition, each trial contained a 35 s long auditory stimulus, composed of 20 familiar triplets (4 triplets, each with 5 repeats, presented in a random order) with no gap between triplets, padded with a 1.5 s silence at the onset and offset of each trial. In the second (unfamiliar) condition, triplets were modified to form unfamiliar sequences the rats had never heard before. Unfamiliar triplets were generated by (1) replacing the middle syllable with another familiar syllable (taken from another triplet), or (2) replacing the middle syllable with an unfamiliar syllable (never heard before), or (3) switching the order of the middle and final syllables, or combinations thereof. Each rat was exposed to 60 familiar stimulus trials and 60 unfamiliar stimulus trials.

After anesthetic induction and ABR recordings, a 4 x 5 mm craniotomy was performed, extending from 0 to 4 mm ventral from the temporal edge and 0 to 5 mm posterior from the Bregma, to expose the right temporal lobe, and the dura matter was carefully removed. Electrophysiological data were recorded by an 8 x 8 rodent ECoG electrode grid at a sampling rate of 24,414 Hz, acquired and amplified by Tucker-Davis Technologies (TDT) RZ2 Bioamp Processor and TDT PZ5 NeuroDigitizer, and controlled by the BrainWare software. While in the present study we recorded responses from both primary and secondary auditory cortical fields, we did not acquire data allowing for channel mapping into separate regions. Auditory stimuli were delivered by

a TDT RZ6 multiprocessor sampling at 48,828 Hz and presented by a custom-made earphone with a flat frequency response (calibrated by a G.R.A.S 46DP-1 microphone). Continuous ECoG data were first band-pass filtered between 0.1-48 Hz using a 6th order two-pass Butterworth filter, and then downsampled to 150 Hz. To emphasize signal components which increase neural response repeatability across stimuli, we epoched the data into segments corresponding to a single triplet presentation and denoised the epoched data using the Dynamic Separation of Sources toolbox (de Cheveigné & Simon, 2008).

## Results and Discussion

First, over multiple training sessions, the performance of all trained animals (n = 4), quantified as the sensitivity index d', was significantly above chance level (see Fig. 7C and Table 1, one sample t-test, all $p < 0.05$). This result indicates that all trained animals could differentiate the familiar triplet stimuli from the unfamiliar control stimuli during the training sessions. The test result showed that the performance (d') of all tested animals (n = 3) was above chance (bootstrap, $p < 0.001$), implying that animals did learn the familiar triplets in one context, and could recognize them when presented among different, unfamiliar triplets in another context. One remaining trained rat was not tested as it did not achieve 70% accuracy in the triplet training sessions which we

had set as a criterion for concluding the training, but the performance during training was still above chance (see Table 1). In the present study, to maintain the novelty of the unfamiliar triplet stimuli used in subsequent electrophysiological recordings, we did not expose rats to the same unfamiliar triplets during the training. While rats could only respond after sound sequence offset, we cannot exclude the possibility that they performed the task by memorizing parts of the triplets, instead of learning entire triplets. However, for the purpose of this study it was sufficient to ensure that the trained and the passive exposure group had extensive opportunity to become familiar with the sound sequences used in subsequent recordings.

To test whether previous findings in humans – namely that familiarity with sound sequences yields spectral peaks in neural activity at rates specific to the duration of those sequences (Ding et al., 2015) – generalize to animal models, we analyzed the ECoG signals in the frequency domain (Fig. 8A). We calculated the Fourier power spectrum values of the continuous ECoG signals and normalized each power spectrum by dividing the power estimate for each frequency point within the 0.25-2.5 Hz range by the sum of all power estimates in this range. We observed that, across all rats, the spectral peaks for both the syllable rate (1.88 Hz) were significantly higher than for the neighboring frequencies (Fig. 8AB; Wilcoxon sign rank tests; $Z_{familiar} = 2.934$, $P_{familiar}$

= 0.003; $Z_{unfamiliar} = 2.934$, $P_{unfamiliar} = 0.003$). While triplet rate peaks were nominally higher relative to neighboring frequencies, reflecting minor peaks in the stimulus spectrum (Fig. 8C), this effect was weak and did not survive Bonferroni correction for multiple comparisons across the four tests ($Z_{familiar} = 2.045$, $P_{familiar} = 0.041$, $Z_{unfamiliar} = 2.401$, $P_{unfamiliar} = 0.016$). Furthermore, no differences in the triplet rate peak were observed between groups (pairwise Wilcoxon rank sum tests between trained, naïve, and passive groups: all $p > 0.2$), suggesting that there were no training-specific spectral signatures in our sample of anesthetized rats. It should be noted that while some earlier studies in humans (Batterink & Paller, 2019) suggested that low-frequency spectral peaks can be observed without attention, other studies demonstrated that diverted attention (Ding et al., 2018) and sleep (Makov et al., 2017) disrupt the neural correlates of acoustic chunking. Therefore, it cannot be ruled out that the animals' state might have influenced the amplitude of spectral peaks, and future studies should test whether performing neural recordings in awake animals could uncover stronger signatures of neural entrainment to the sequence presentation rate.

Next, we turned to the analysis of neural signals in the time domain. We applied a principal component analysis to the denoised data after concatenating data across all animals and trials, and retained the first component (explaining 52.96% of the variance).

Then, we segmented the data into epochs, separately for each animal, trial, and triplet. Based on the segmented data, we calculated the average root-mean-square (RMS) amplitude over 20 triplets, separately for each trial and time point. Here we found that the response waveforms of ECoG responses were quite diverse across rats, and there was no consistent difference in responses to familiar and novel stimuli between trained, passive, and naïve animals. Specifically, a repeated-measures ANOVA was conducted to compare the single-trial RMS values of time courses between conditions (familiar vs. unfamiliar) and groups (trained vs. naïve). In the ANOVA, neither the main effect of condition (all $p > 0.05$, FDR-corrected across time points) nor a significant interaction between group and condition was observed (all $p > 0.05$, FDR-corrected). However, we observed a significant main group effect at all three syllables (all $F_{2,1298} > 10.826$, all $p < 0.05$, FDR-corrected), indicating that acoustic training had modulated the overall neural responses to sound stimuli. Furthermore, we compared the RMS amplitude curves separately for each animal between the two conditions (familiar vs. unfamiliar; Fig. 9) and found significant differences at the second and third syllable positions in all animals except one (t-test, all $p < 0.05$, FDR-corrected across time points). Still, the direction of these differences was not consistent across syllables and rats (as evident from the lack of the main effect of condition in the repeated-measures ANOVA), indicating that average neural responses to familiar and unfamiliar stimuli were

heterogeneous in our sample (An et al., 2021). Thus, unlike previous findings (Sanders et al., 2002, 2009), showing that a larger N1 amplitude of the EEG response at the first-syllable position in human subjects after familiarization may be a neural correlate of auditory chunking, in the present study we did not observe consistent amplitude differences or peaks in the time course analysis. While this suggests that the neural responses averaged over the entire AC region are not a sensitive measure of prior experience with familiar sequences, it does raise the possibility that sequence familiarity might instead affect more fine-grained population responses, accessible through multivariate analyses.

To determine whether stimuli can be decoded from neural activity, and whether stimulus familiarity affects decoding, we quantified the relative multivariate Mahalanobis distance (dissimilarity; (Ledoit & Wolf, 2004)) between spatial patterns of neural responses to different familiar and unfamiliar stimuli. In the first step, we calculated the RMS over time points of the ECoG amplitude for each channel, stimulus (i.e., each triplet), syllable, and rat. Then, per stimulus type (familiar vs. unfamiliar), syllable, and rat, we used a leave-one-out cross-validation method to calculate the single-trial Mahalanobis distance values between the vector of RMS amplitudes concatenated across channels for a particular trial, and four vectors of average RMS

amplitudes (averaged across the remaining trials, separately for each of the four triplets) concatenated across channels. The single-trial decoding estimate was calculated as a difference between (1) the dissimilarity between RMS amplitudes observed in this trial vs. other trials of the same stimulus, and (2) the average dissimilarity between RMS amplitudes observed in this trial vs. other stimuli. These single-trial decoding estimates were averaged across trials, separately for each stimulus type (familiar vs. unfamiliar), syllable, and rat. Finally, per syllable and rat, we calculated a ratio between the averaged decoding estimates for familiar vs. unfamiliar stimuli. These ratios were compared between trained, passive, and naïve rats, separately for each syllable.

As expected, no difference between trained, passive, and naïve animals was observed for the first syllable (permutation tests; trained vs. naïve: $p = 0.301$; trained vs. passive: $p = 0.192$; passive v. naïve: $p = 0.337$; Fig. 10), since the first syllables were physically identical in both types of triplets. At the second syllable, while we did not find evidence for a consistent, significant difference between groups (trained vs. naïve: $p = 0.190$; trained vs. passive: $p = 0.300$; passive v. naïve: $p = 0.080$; Fig. 10), one of the four trained rats showed robust improvement in decoding for familiar vs. unfamiliar stimuli. Crucially, we found a robust difference between groups at the third syllable position, where the decoding values in the trained group were significantly higher than in the

remaining groups group (trained vs. naïve: $p = 0.010$; trained vs. passive: $p = 0.002$; passive v. naïve: $p = 0.240$; Fig. 10).

Consistent with our hypothesis and with previous studies in the visual modality (Gavornik & Bear, 2014), all trained animals were more sensitive to the familiar stimuli than to the unfamiliar stimuli (decoding ratio, including 95% confidence intervals, above 1 for all trained animals; Fig. 10). Our results are also consistent with a previous study in humans (Batterink et al., 2015), in which familiarity effects were strongest on the late/final sequence elements, and were specifically improved by previous extensive training, but not by passive exposure. Importantly, in our study the decoding boost was observed in neural activity recorded during post-training anesthesia. While one previous study in cats has identified rapid effects of passive exposure to visual sequences (movies with natural scenes) on cortical activity under anesthesia (Yao et al., 2007), most studies in animal models used extensive behavioral training and recordings in awake animals to identify the behavioral and/or neural correlates of sequence processing (Bale et al., 2017; Gavornik & Bear, 2014; Homann et al., 2017; Murphy et al., 2008). However, given a recent study based on single-neuron recordings in the auditory cortex of awake mice (Libby & Buschman, 2021), in which neural signatures of passive sequence learning were shown to gradually increase over several days, it

cannot be ruled out that decoding would also have been observed in our passive group after more extensive exposure. Nevertheless, the effect of prior training on shaping the neural responses to familiar stimuli under anesthesia is consistent with earlier findings in rats that training-induced receptive field plasticity shifts the sensitivity of neuronal populations in the auditory cortex toward the reinforced sound frequency (David et al., 2012; Fritz et al., 2005). Here, we show that learning experience boosts the ability to decode stimulus information from neural activity also in the context of learning temporally extended sequences.

In summary, we show that training experience can improve the neural sensitivity to sequences in rodents, although the neural correlates typically observed in humans (low-frequency entrainment to sequence presentation rate) are not detectable in the neural activity of anesthetized rodents. Instead, we show that behavioral training leads to improvements in decoding stimulus-related information from the spatial pattern of neural activity in the auditory cortex, even under anesthesia. Future studies should test the behavioral relevance of these signals by relating the neural activity to behavioral responses in awake and behaving animals.

**Figure 7**

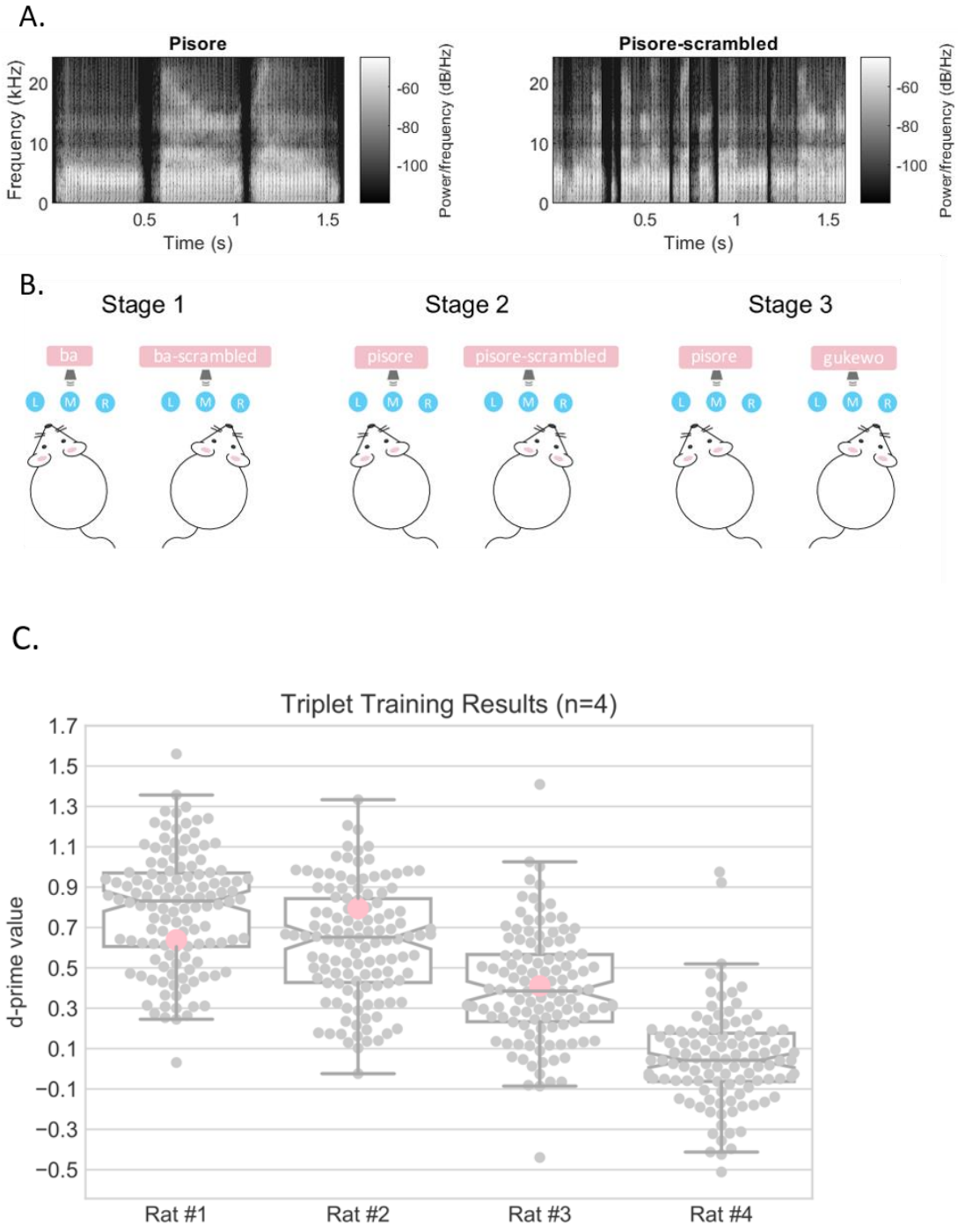


**Figure 7. Behavioral stimuli and training paradigm.** (A) Spectrograms of two example stimuli (/pisore/ and its scrambled equivalent). (B) Two-alternative forced choice training paradigm. There were three training stages in total. In training stage 1,

there was an additional spatial cue for helping rats to recognize the syllables. This spatial cue was gradually removed as rats reached a threshold of 75% accuracy in discriminating familiar (i.e. /ba/→lick left side) and unfamiliar stimuli (i.e. /ba/-scrambled → lick right side). In training stages 2 and 3, rats were conditioned to discriminate familiar (i.e. /pisore/ → lick left side) and unfamiliar (i.e. /pisore/-scrambled or /gukewo/ → lick right side) triplets rather than single syllables. (C) Behavioral performance of individual rats during training. The grey dots denote the training results of each training session (d-prime values) and the pink dots represent the test results (d-prime values) in the testing session. Performance of all rats was above chance level (bootstrapped, all $p<0.05$), and three out of four rats which approached 70% accuracy in the triplet learning stage also performed significantly above-chance in the test session (chance level: $d'=0$, bootstrapped, $p<0.05$).

**Figure 8**

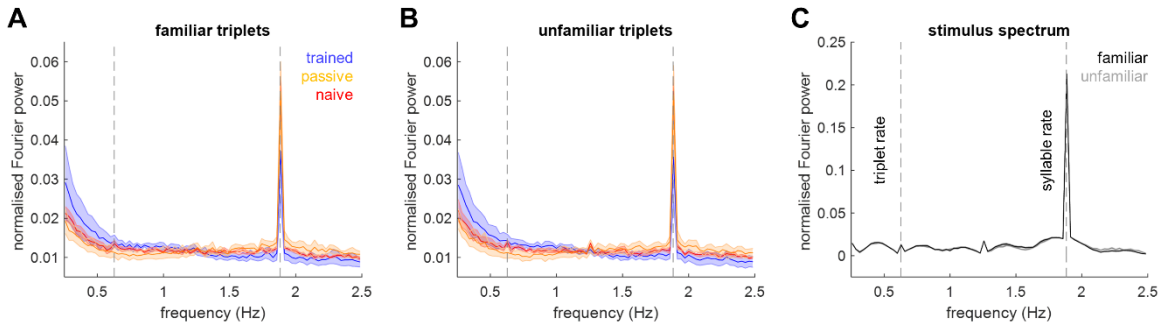


**Figure 8. Frequency-domain neural activity.** (A) Normalized power spectrum values calculated for familiar triplets in the trained group (n=4; blue), passive group (n=3, orange) and naïve group (n=4, red), showing a robust peak at the syllable rate but no robust peak at the triplet rate. No significant differences were observed between groups. Shaded areas denote SEM across rats. (B) Average power spectra for unfamiliar triplets. Legend as in (A). (C) Average power spectra based on stimulus envelope. Black: familiar stimuli; grey: unfamiliar stimuli; shaded areas denote SEM across trials.

**Figure 9**

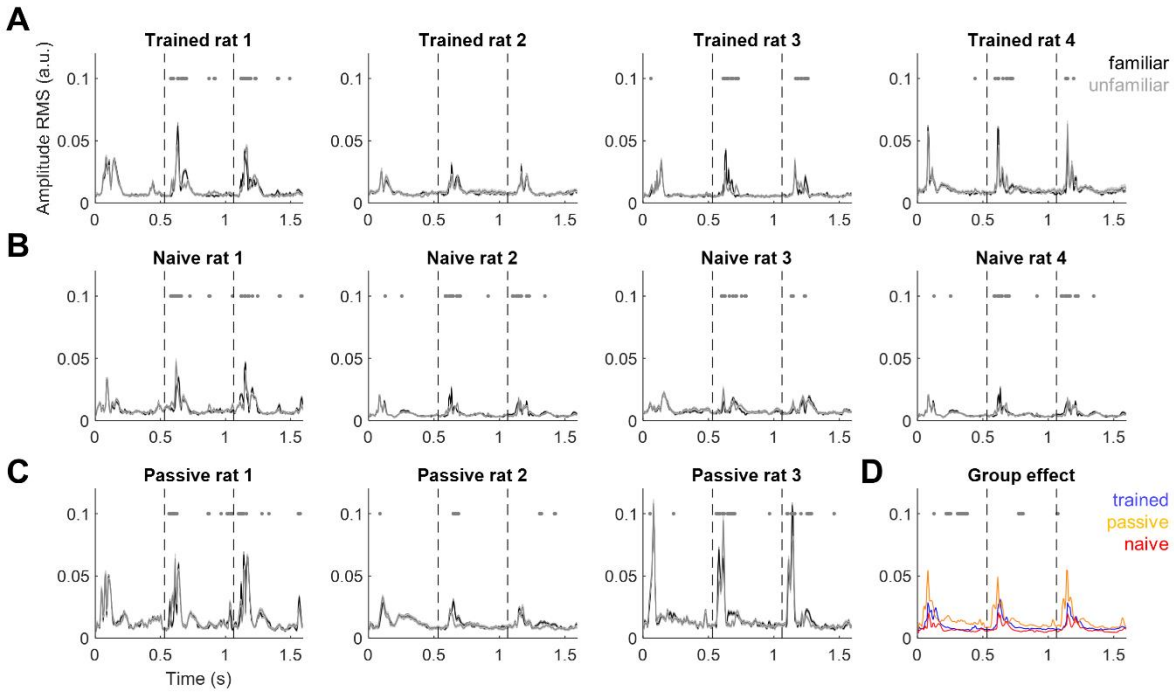


**Figure 9. Time-domain neural activity.** (A-C) Blue/red lines denote the (RMS) amplitude over triplets of the ECoG response to familiar/unfamiliar sequences for each individual rat from the trained group (A), naïve group (B), and passive group (C). Grey dots show the significance of the main effect of stimulus familiarity per time point of the RMS amplitude (t-test, $P<0.05$, FDR corrected). (D) RMS time courses averaged across individual rats, per group. Grey dots show a significant main effect of group (ANOVA, $P<0.05$, FDR corrected).

**Figure 10**

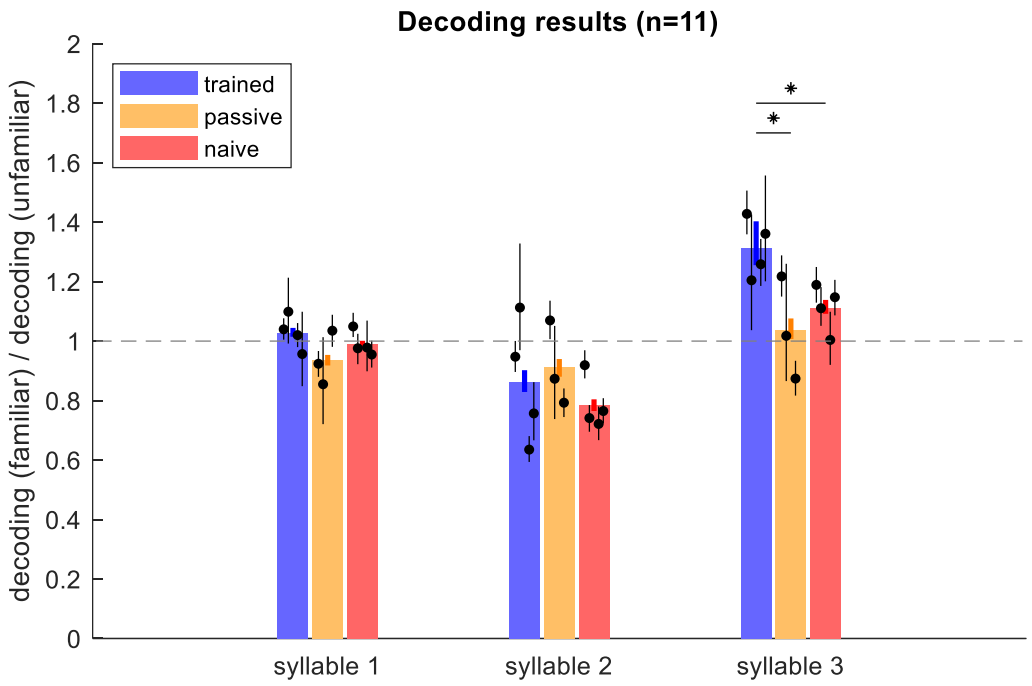


**Figure 10. Multivariate analysis.** The blue bars show the decoding index (the ratio of decoding familiar stimuli to decoding unfamiliar stimuli) for trained rats whereas the orange/red bars indicate the decoding index for the passive/naïve groups respectively. Thick blue/orange/red lines indicate the 95% confidence interval (CI) range of decoding indices at the group level. Each black dot represents the respective decoding index per rat, and each black line stands for the 95% CI of individual rats' decoding indices. Asterisks denote significant differences between groups ($p < 0.05$).

**Table 1**

Behavioral Results in Triplet Discrimination Training Stage

| | Total Session No. | Trial No. /Session (Mean±SD) | Accuracy (%) (Mean±SD) | Significance (P-values) | T-values | df | Accuracy in Test Session (%)/Trial No. |
|---|---|---|---|---|---|---|---|
| **Rat #1** | 127 | 287.42±97.02 | 64.14±5.39 | P<0.0001 | 29.54 | 126 | 62.54/289 |
| **Rat #2** | 123 | 301.08±93.77 | 61.88±5.26 | P<0.0001 | 25.02 | 122 | 64.88/197 |
| **Rat #3** | 118 | 235.34±91.54 | 57.03±6.05 | P<0.0001 | 12.62 | 117 | 56.58/239 |
| **Rat #4** | 112 | 260.02±89.33 | 50.86±4.13 | P=0.029 | 2.02 | 111 | Nil |

**Table 1.** Statistical information. Behavioral training information and results of all animals.

## References

An, H.J., Auksztulewicz, R., Kang, H.J., Schnupp, J.W.H., 2021. Cortical mapping of mismatch responses to independent acoustic features. Hear. Res. 399, 107894.

Bale, M.R., Bitzidou, M., Pitas, A., Brebner, L.S., Khazim, L., Anagnou, S.T., Stevenson, C.D., Maravall, M., 2017. Learning and recognition of tactile temporal sequences by mice and humans. Elife 6, e27333.

Batterink, L.J., Paller, K.A., 2019. Statistical learning of speech regularities can occur outside the focus of attention. Cortex 115, 56-71.

Batterink, L.J., Reber, P.J., Paller, K.A., 2015. Functional differences between statistical learning with and without explicit training. Learn. Mem. 22(11), 544-556.

David, S.V., Fritz, J.B., Shamma, S.A., 2012. Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc. Natl. Acad. Sci. U.S.A. 109(6), 2144-2149.

de Cheveigné, A., Simon, J.Z., 2008. Denoising based on spatial filtering. J. Neurosci. Methods. 171(2), 331-339.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., Pallier, C., 2015. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. Neuron 88(1), 2-19.

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., Poeppel, D., 2017. Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG).

Front. Hum. Neurosci. 11, 481.

Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2015. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19(1), 158-164.

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., Zhang, J., 2018. Attention is required for knowledge-based sequential grouping: Insights from the integration of syllables into words. J. Neurosci. 38(5), 1178-1188.

Edeline, J.M., Weinberger, N.M., 1993. Receptive field plasticity in the auditory cortex during frequency discrimination training: selective retuning independent of task difficulty. Behav. Neurosci. 107(1), 82.

Fritz, J., Elhilali, M., Shamma, S., 2005. Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. Hear. Res. 206(1-2), 159-176.

Gavornik, J.P., Bear, M.F., 2014. Learned spatiotemporal sequence recognition and prediction in primary visual cortex. Nat. Neurosci. 17(5), 732-737.

Homann, J., Koay, S.A., Glidden, A., Tank, D., Berry, M., 2017. Predictive Coding of Novel versus Familiar Stimuli in the Primary Visual Cortex. BioRxiv 197608.

Ives, D.T., Smith, D.R.R., Patterson, R.D., 2005. Discrimination of speaker size from syllable phrases. J. Acoust. Soc. Am. 118(6), 3816-3822.

Jiang, X., Long, T., Cao, W., Li, J., Dehaene, S., Wang, L., 2018. Production of Supra-regular Spatial Sequences by Macaque Monkeys. Curr. Biol. 28(12), 1851-1859.

Kawahara, H., 2006. STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. Acoust. Sci. Technol. 27(6), 349-353.

Kelly, J.B., Masterton, B., 1977. Auditory sensitivity of the albino rat. J. Comp. Physiol. Psychol. 91(4), 930.

Kikuchi, Y., Attaheri, A., Wilson, B., Rhone, A.E., Nourski, K.V., Gander, P.E., Kovach, C.K., Kawasaki, H., Griffiths, T.D., Howard III, M.A. and Petkov, C.I., 2017. Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. PLoS Biol. 15(4), e2000219.

Ledoit, O., Wolf, M., 2004. Honey, I shrunk the sample covariance matrix. J. Portf. Manag. 30(4), 110-119.

Li, K., Chan, C.H.K., Rajendran, V.G., Meng, Q., Rosskothen-Kuhl, N., Schnupp, J.W.H., 2019. Microsecond sensitivity to envelope interaural time differences in rats. J. Acoust. Soc. Am. 145(5), EL341-EL347.

Libby, A., Buschman, T.J., 2021. Rotational dynamics reduce interference between sensory and memory representations. Nat. Neurosci. 24(5), 715-726.

Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., Golumbic, E.Z., 2017. Sleep disrupts high-level speech parsing despite significant basic auditory processing. J.

Neurosci. 37(32), 7772-7781.

Murphy, R.A., Mondragón, E., Murphy, V.A., 2008. Rule learning by rats. Science 319(5871), 1849-1851.

Romberg, A.R., Saffran, J.R., 2010. Statistical learning and language acquisition. Wiley Interdiscip. Rev. Cogn. Sci. 1(6), 906-914.

Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. Science 274(5294), 1926-1928.

Sanders, L.D., Ameral, V., Sayles, K., 2009. Event-related potentials index segmentation of nonsense sounds. Neuropsychologia 47(4), 1183-1186.

Sanders, L.D., Newport, E.L., Neville, H.J., 2002. Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. Nat. Neurosci. 5(7), 700-703.

Toro, J.M., Trobalón, J.B., 2005. Statistical computations over a speech stream in a rodent. Percept. Psychophys. 67(5), 867-875.

Wilson, B., Marslen-Wilson, W.D., Petkov, C.I., 2017. Conserved Sequence Processing in Primate Frontal Cortex. Trends Neurosci. 40(2), 72-82.

Wilson, B., Slater, H., Kikuchi, Y., Milne, A.E., Marslen-Wilson, W.D., Smith, K., Petkov, C.I., 2013. Auditory artificial grammar learning in Macaque and Marmoset

monkeys. J. Neurosci. 33(48), 18825-18835.

Yao, H., Shi, L., Han, F., Gao, H., Dan, Y., 2007. Rapid learning in cortical coding of visual scenes. Nat. Neurosci. 10(6), 772-778.

# Chapter 3. Deviant Processing of Complex Sounds in Mouse Auditory Cortex

## Introduction

Deviance detection is a fundamental building block of cognition as it allows for learning and optimizing internal models of the world (Auksztulewicz & Friston, 2016). The mammalian auditory system encodes the same stimuli differently, depending on the context in which they are presented, such as the order of the stimulus in a sequence (Herrmann et al., 2015; Perez-Gonzalez et al., 2005; Yaron et al., 2012), or whether the stimulus is expected or unexpected based on the statistical regularities in the auditory inputs (Dehaene et al., 2015). The classical oddball paradigm has been extensively used for studying the neural mechanisms of deviance detection (Naatanen et al., 2007; Ulanovsky et al., 2003). It involves presenting a sequence of identical repeating stimuli (‘standards’), which as a result have a high probability, and oddball stimuli (‘deviants’), which have a low probability (Cowan et al., 1993). The difference between neural responses to the deviant stimulus and standard stimulus is known as a mismatch response, and accumulating evidence suggests that MMRs can reflect prediction error signals, rather than mere release from adaptation to the standard (Auksztulewicz & Friston, 2015; Carbajal & Malmierca, 2018; Friston, 2005; Garrido et al., 2009; May &

Tiitinen, 2010).

Both humans (Chennu et al., 2016; Wacongne et al., 2011) and macaques (Uhrig et al., 2014) (Wang et al., 2015) can detect deviants not only based on local probabilities of single stimuli (i.e., in sequences of the type A|B, where a particular standard stimulus A has a high probability and deviant stimulus B has lower probability), but also based on more complex patterns multiple elements, which may be "chunked" or grouped perceptually according to observed transition probabilities, (consider phonemes grouped into syllables and syllables grouped into streams of vocalizations as an example). In such contexts, whether individual sounds are expected or not may depend either on their "local" or their "global" context (Chao et al., 2018). Similarly, in humans, the auditory cortex elicits MMR also based on unexpected repetitions (AB|AA) and unexpected omissions (AB|A_), which suggests that humans chunk the standard pair (AB) and encode the prediction error to chunk violations (Chouiter et al., 2015; Todorovic & de Lange, 2012; Todorovic et al., 2011). While rodents can detect single deviants (Ulanovsky et al., 2003; Yaron et al., 2012) and change in acoustic features (An et al., 2021; Yang et al., 2021), to the best of our knowledge there is no direct evidence for MMR signaling in complex sound sequences, although rats can perform sequence chunking if extensively trained (Luo et al., 2021) or if the sequences are

repetitive enough (Cappotto et al., 2021).

In rodents, an extensive literature on mismatch responses to single stimuli suggests a hierarchy of regions, such that secondary areas encode deviance detection to a larger extent than primary areas (Parras et al., 2021; Parras et al., 2017). In addition, an earlier study (Chen et al., 2015) using two-photon guided patch recording in the primary auditory cortex of anesthetized mice suggested that the early component (0-100 ms after tone onset) of the neural response to a deviant sound may reflect neural adaptation, while the late component (200 – 400 ms after onset) may reflect deviance detection. However, it remains elusive whether the primary auditory cortex, or higher order regions, can also encode deviance in more complex stimulus patterns.

In this study, we used continuous, repetitive trains of syllable pairs consisting of artificial syllables (‘pe bi’) as high probability standard stimuli, and manipulated the elements of the pair to obtain three different types of low probability deviants: (1) substitution deviants (‘pe bi’→‘pe da’ or ‘pe bi’→’da bi’), (2) transposition deviants (‘pe bi’→‘bi pe’), and (3) omission deviants (‘pe bi’→‘pe _’). Note that the silent interval within a syllable pair (0.145 s) was much shorter than the interval between syllable pairs (2.35 s), which encourages perceptual grouping of the syllable stream into

disyllabic “chunks”. This grouping sets up a context in which deviations from the expected ‘pe bi’ can be thought of as occurring on a local, within-syllable-pair level (substitutions or omissions) or on a more global level involving the whole pair (transpositions). As a control condition, we also presented streams of the same syllable pairs (pe bi’,‘pe da’, ‘da bi’, ‘bi pe’ and ‘pe _’) but with equal probability (see Fig. 11a). We performed wide-field calcium recordings (see Fig. 11b), imaging from multiple auditory cortical areas of awake mice. We focused on testing whether cortical activity in mouse auditory cortex, as revealed by Ca++ imaging, is sensitive to sequence violations, and whether the character of this sensitivity differs for within-pair vs more global, whole-pair violations. Further, we investigated whether mismatch responses to different deviant types differ between primary and higher-order regions.

**Figure 11**

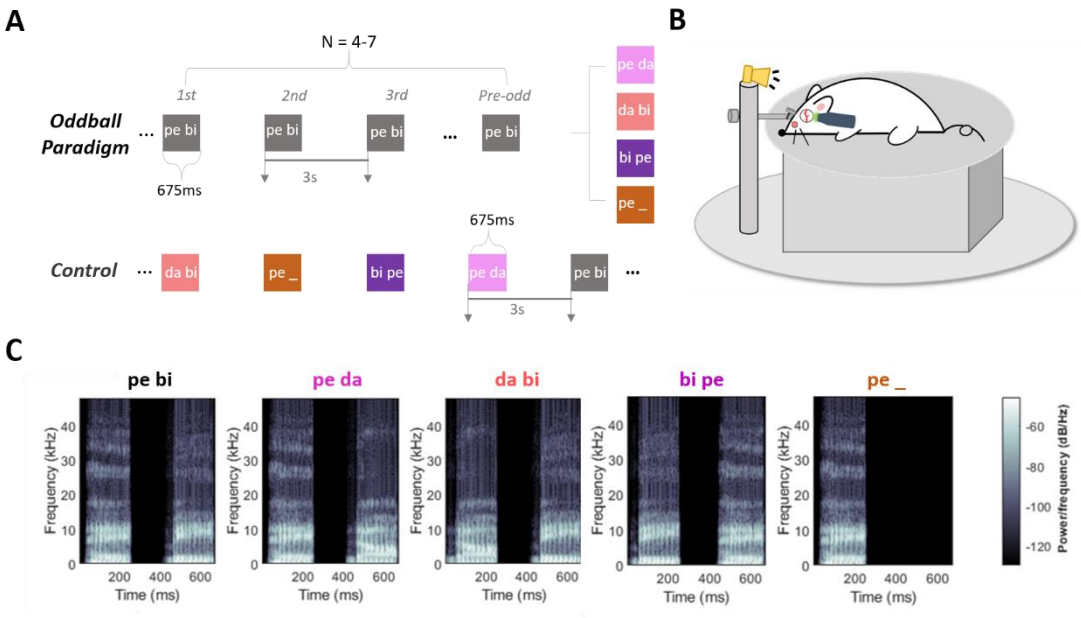

Figure 11. (A) Schematic of auditory stimuli paradigm. (B) Customized experimental set-up for awake recording. Head-fixed Gcamp6s transgenic mice passively listened to auditory stimuli during wide-field microscopy imaging of the auditory cortex. (C) Spectrograms of auditory stimuli.

## Results

### *Parcellation of regions into primary and higher-order auditory cortex*

**Figure 12**

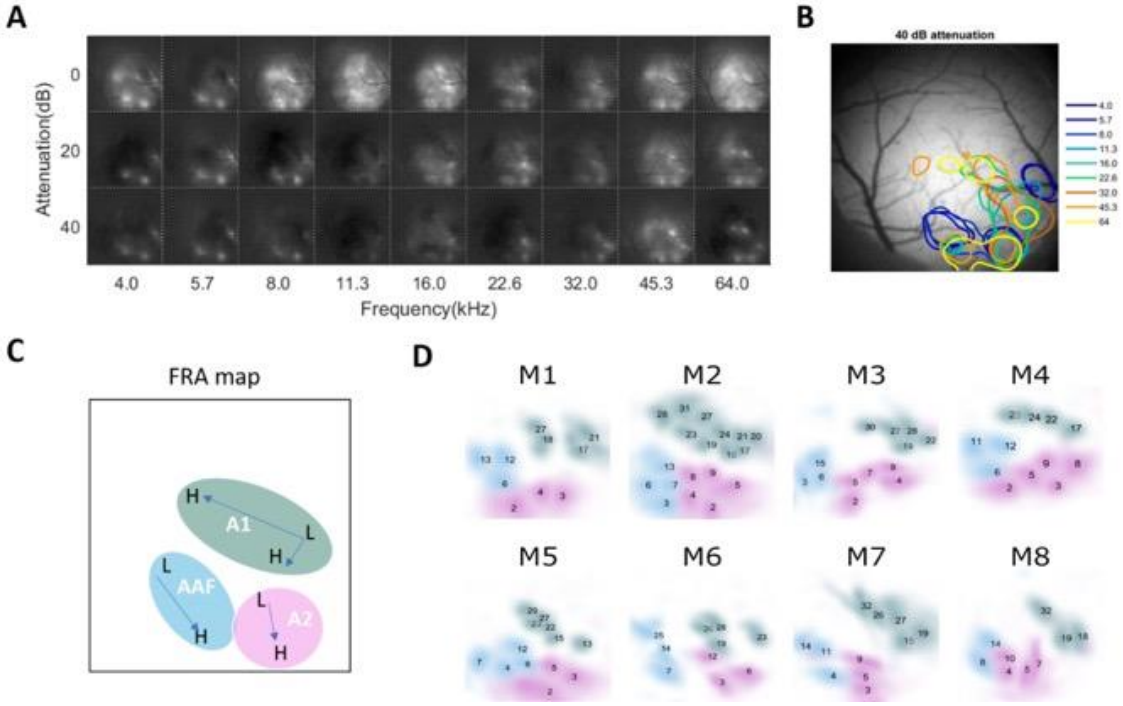


Figure 12. (A) Example of WF calcium signal to pure tones of different frequencies and sound levels of pure tones in one animal. Each image shows the averaged evoked response within 200-500ms after tone onset. (B) Example of tonotopy map at 40 dB attenuation in one animal. Different frequencies (kHz) of pure tone are color-coded, and each contour line illustrates the response area to the corresponding frequency. (C) Schematic of a tonotopic FRA map. Arrows represent gradients of low to high frequency preference. Distinct auditory fields (A1, AAF, A2) marked in different colors.

(D) Tonotopic FRA maps of 8 animals. Color indicates the corresponding auditory fields.

We performed imaging of the left auditory cortex of ten awake adult mice, but excluded two animals due to poor data quality caused by cloudy cranial windows. Before characterizing the cortical responses to the syllable stimulus trains just described, we first identified the spatial location of different auditory cortical fields by imaging the frequency response areas (FRA) to different frequencies of pure tones (4-64 kHz in half octave steps) at different attenuation levels (0 dB, 20 dB, 40 dB). We found a clear tonotopy map for all analyzed subjects (N=8). The fluorescence amplitudes increased most at the 0 dB attenuation level, while the tonotopic map was much clearer at 40 dB attenuation. Based on the relative positions and the corresponding tonotopic gradients of different areas in the FRA map (Fig. 12C), we subdivided the imaged auditory cortical areas into A1, AAF, and A2. The tonotopic organization was reliable across all animals (Figure 12D) and was consistent with the previous study using the same type of transgenic animals (Liu et al., 2019; Romero et al., 2020). Finally, we selected the components for each area based on the tonotopic map from the ROIs generated by an autoencoder analysis as described in (Liu et al., 2019).

***Neuronal adaptation to repetition of auditory stimuli in the Oddball condition***

**Figure 13**

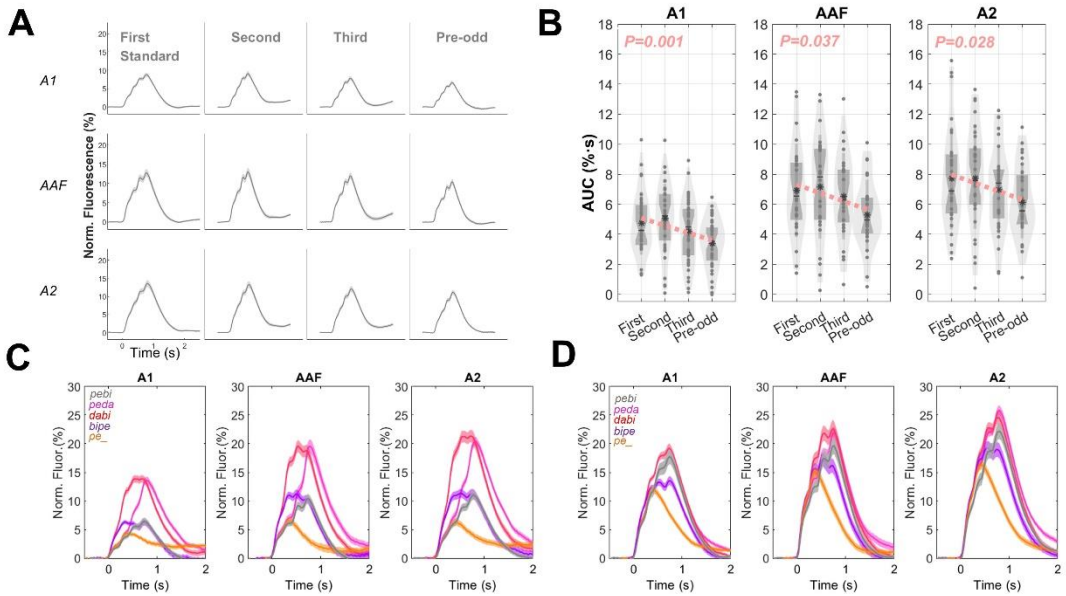


Figure 13. (A)Neural adaptation to subsequent repetitions of standard stimuli. (B)Linear regression quantifying repetition suppression. (C) Calcium response to deviants and standards in oddball condition. (D) Calcium response to the same stimulus types in the control condition.

First, we quantified neuronal adaptation by analysing the responses to subsequent standard syllables. We observed that the fluorescence of the response to standard stimuli decreased from the first standard to the last repeat (immediately preceding the oddball stimulus; “pre-odd”) A1, AAF and A2 (See Fig. 13A). To test whether there is a neural adaptation effect, per area we performed linear regression on the area under the curve (AUC) of the fluorescence in response to the first three standard repeats and the last repeat. We identified a significant decrease of the AUC values from the first repeat to the last repeat in all areas (A1: $F(2,150) = 10.7$, $P = 0.001$; AAF: $F(2,102) = 4.46$, $P =$

0.037; A2: F(2,118) = 4.93, P = 0.028; see Fig. 13B).

Then, to test whether the syllable pairs differed significantly in how strongly they activate cortex per se, independently of oddball or frequency effects, we performed two-way ANOVA on AUC values of fluorescence. We tested for differences between the four stimulus types (the standard and 3 deviants) and auditory areas (A1, AAF, A2), separately for the oddball condition and control condition.

In the control condition, there was a significant main effect of stimulus type ($F_{(4,455)}$ = 36.06, P < 0.001) and areas ($F_{(2,455)}$ =36.77, P < 0.001). However, the interaction between stimulus type and areas was not significant ($F_{(8,455)}$ = 0.65, P = 0.737).

In the oddball condition, the main effect of the stimulus type was also significant ($F_{(4,737)}$ = 158.58, P < 0.001) as was the main effect of auditory areas ($F_{(2,737)}$ = 80.37, P < 0.001). In addition, we found a significant interaction between stimulus types and auditory areas ($F_{(8,737)}$ = 3.17, P = 0.002). In the post-hoc comparison in the oddball condition, we observed significant differences between deviants (averaged over type) and standards (A1: $t_{37}$ = 11.758, P<0.0001, AAF: $t_{25}$ = 10.792, P<0.0001, A2: $t_{29}$ = 12.560, P<0.0001). Then we compared the overall response between areas, and we found the overall response in A2 and AAF (post hoc, both p<0.0001) to be significantly stronger than in A1, and there is no significant difference between AAF and A2 (post

hoc, P=0.160).

Thus, these results suggest that, in the oddball condition, stimulus frequency modulates the responses to stimuli, and this modulation might differ between auditory region. Such effect was not found in the control condition and we would not include the control condition for further comparison.

### *Wide field responses to complex deviants*

**Figure 14**

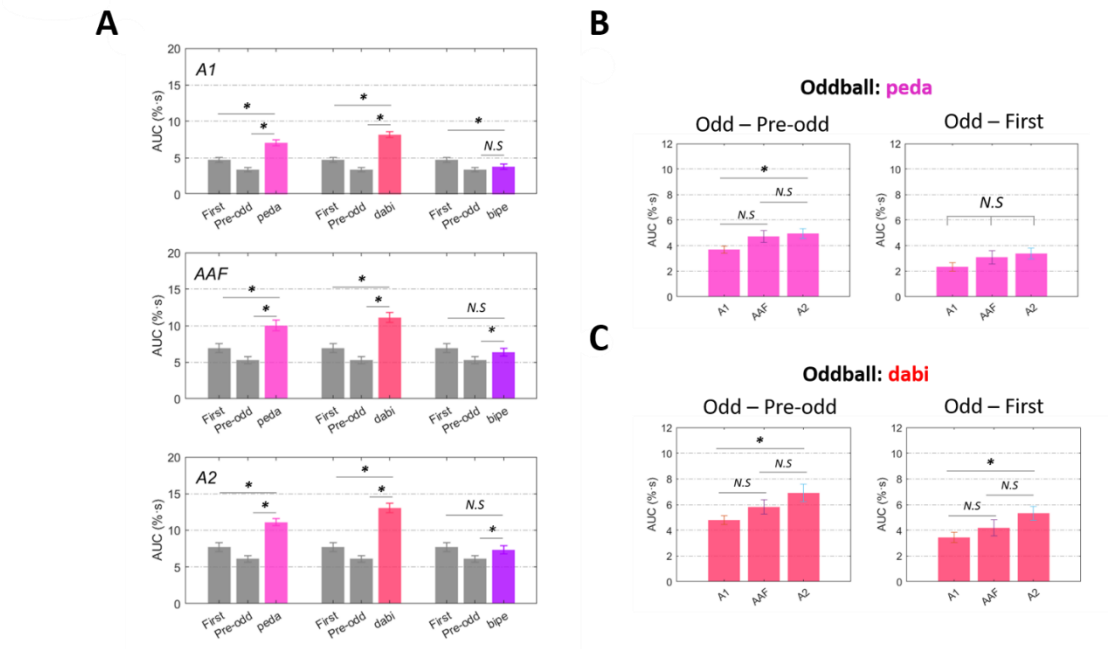


Figure 14. (A) Comparison of calcium response to the first standard stimuli, pre-odd standard stimuli and oddball stimuli across different deviant types and distinct auditory fields. (B) Comparison of mismatch response to the oddball ($AUC_{oddball}$) and the responses to the standard preceding the oddball ($AUC_{pre\text{-}odd}$) or to the first standard following the oddball ($AUC_{first}$) between areas in oddball condition (substitution

oddball, ‘pe da’). (C) Comparison of mismatch response to the oddball ($AUC_{oddball}$) and the responses to the standard preceding the oddball ($AUC_{pre\text{-}odd}$) or to the first standard following the oddball ($AUC_{first}$) between areas in oddball condition (substitution oddball, ‘da bi’).

To estimate the mismatch responses evoked by different types of deviant stimuli, first we compared the AUC of fluorescence to the substitution (‘pe da’, ‘da bi’) and transposition (‘bi pe’) deviant stimuli against the pre-odd standard stimuli (‘pe bi’) in oddball blocks. This analysis showed that, for deviants containing a substitution of the second element of the pair (‘pe da’), the AUC was significantly larger for oddball stimuli compared with the pre-odd standard stimuli. This was true in all auditory fields (A1 area, paired t-test: $t_{37} = 13.672$, $P_{peda} < 0.001$; AAF area, paired t-test: $t_{25} = 9.832$, $P_{peda} < 0.001$; A2 area, paired t-test: $t_{29} = 13.411$, $P_{peda} < 0.001$; FDR corrected). Similarly, for deviants containing a substitution of the first pair element (‘da bi’), we found the same pattern of results in all areas (A1 area, paired t-test: $t_{37} = 14.240$, $P_{dabi} < 0.001$; AAF area, paired t-test: $t_{25} = 10.126$, $P_{dabi} < 0.001$; A2 area, paired t-test: $t_{29} = 10.342$, $P_{dabi} < 0.001$; FDR corrected). Interestingly, for the transposition condition (‘bi pe’), we found significant differences between the AUC of the deviants and the pre-odd standard in the AAF and A2 areas (AAF area: paired t-test: $t_{25} = 3.031$, $P_{bipe} = 0.006$; A2 area: paired t-test: $t_{29} = 3.224$, $P_{bipe} = 0.003$; FDR corrected), but not in the A1 area

(paired t-test: $P_{bipe} > 0.05$; FDR corrected).

In the above analysis, we compared the responses to a deviant stimulus vs. to an adapted pre-odd standard. Next, we sought to quantify the mismatch response after taking adaptation into account. To this end, we compared the AUC of fluorescence to the deviant stimuli ('pe da', 'da bi', 'bi pe') against the first standard stimuli ('pe bi') in oddball blocks. The results showed that for substitution deviants ('pe da' and 'da bi'), the AUC of response to deviants was significantly larger than that to the first standard stimuli in all auditory fields (A1 area, paired t-test: $t_{37} = 6.916$, $P_{peda} < 0.001$; AAF area, paired t-test: $t_{25} = 5.836$, $P_{peda} < 0.001$; A2 area, paired t-test: $t_{29} = 7.990$, $P_{peda} < 0.001$; A1 area, paired t-test: $t_{37} = 8.716$, $P_{dabi} < 0.001$; AAF area, paired t-test: $t_{25} = 6.798$, $P_{dabi} < 0.001$; A2 area, paired t-test: $t_{29} = 9.646$, $P_{dabi} < 0.001$; FDR corrected for all P values). On the other hand, for the transposition deviant ('bi pe'), we found that there was a significant difference between the AUC of the deviants vs. the first standard in the A1 area (paired t-test: $t_{37} = -2.862$, $P_{bipe} = 0.006$; FDR corrected), but not in the AAF and A2 areas (AAF area, paired t-test: $P_{bipe} > 0.05$; A2 area, paired t-test: $P_{bipe} > 0.05$; FDR corrected).

Having shown that there are significant mismatch responses (difference between

$AUC_{oddball}$ and $AUC_{first,}$ all P <0.001, FDR corrected, see Fig. 14A) to substitution deviants in all fields, and to transposition deviants in a subset of fields (AAF and A2 , $P_{AAF}$=0.006,$P_{A2}$=0.003, FDR corrected, see Fig. 14A), we then tested for differences in mismatch responses between fields. To this end, for each area and deviant type we subtracted responses to standards from responses to deviants, and entered the difference values into rank-sum tests between areas. We found there was a significant difference between A1 and A2 in mismatch responses to substitution deviants compared with the pre-odd standards (rank-sum test, $Z_{peda}$ = 2.773, $P_{peda}$ = 0.006, $Z_{dabi}$ = 3.005, $P_{dabi}$ = 0.002, FDR corrected,see Fig.14C), but not between A1 and AAF (rank-sum test, all P > 0.05, FDR corrected), or between AAF and A2 (rank-sum test, all P > 0.05, FDR corrected). However, when analyzing mismatch responses relative to the first standard, we only identified a significant difference between A1 and A2 in mismatch responses to deviants based on substituting the first element of the pair (‘da bi’; mismatch response significantly stronger in A2, rank-sum test, Z = 2.612, P = 0.009, FDR corrected), but not between other pairs of regions, and not for other types of deviants (all other P > 0.05).

We observed no significant omission responses in this study. The AUC of fluorescence to the omission deviant stimuli was found to be smaller than to the standards in all

auditory cortical fields (deviant vs pre-odd: paired t-test, $t_{37}$ = 2.929, $P_{A1}$=0.006, $t_{25}$=5.919, $P_{AAF}$<0.001, $t_{29}$=9.807, $P_{A2}$<0.001, FDR corrected; deviant vs fist standard: paired t-test, A1($t_{37}$ = 10.065), AAF($t_{25}$=9.844), A2($t_{29}$=10.175), all P <0.001, FDR corrected, see Fig. 15C).

Beyond quantifying mismatch responses averaged over time (using AUC of fluorescence values), given previous reports that neural activity at different latencies might be functionally dissociable (Chen et al., 2015), we also aimed at characterizing mismatch responses with a higher temporal resolution. To this end, we calculated the first temporal derivative values for the calcium responses traces (fluorescence) and subjected them to further analysis (see Fig. 15A). In the temporal derivative time series, we observed four distinct response peaks (i.e., an early and a late response to each stimulus in a pair), based on which we specified four distinct time windows for the sub-syllabic responses: the early peak of the first syllable (0-133 ms after the first stimulus onset), the late peak of the first syllable (166-300 ms after the first stimulus onset), the early peak of the second syllable (0-133 ms after the second stimulus onset) and the late peak of the second syllable (166-300 ms after the second stimulus onset).

**Figure 15**

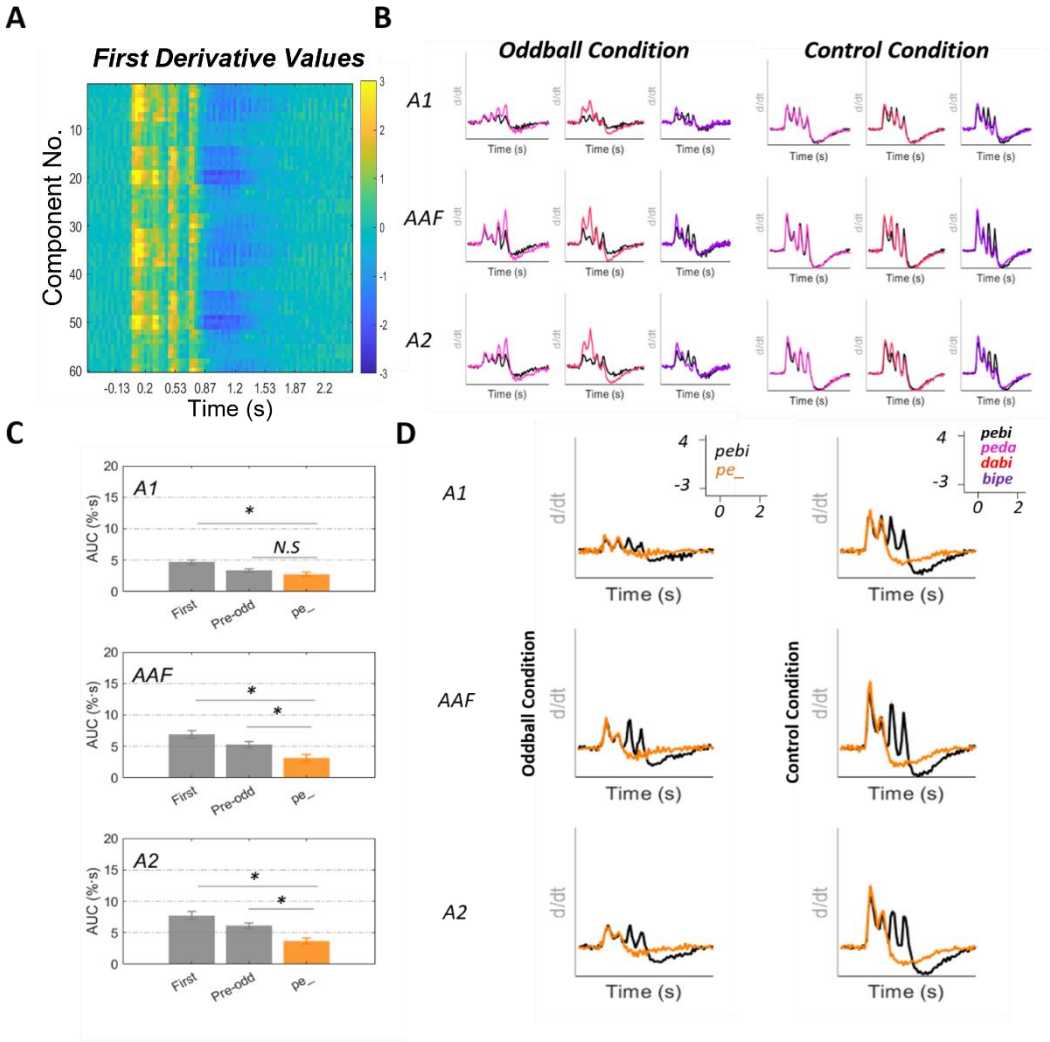


Figure 15. (A) The first derivatives of the traces of standard stimuli from all components in the control condition. Four peaks of derivative values in different time windows were observed. (B)The first derivatives of the traces (dark: standard, color-coded: deviants) from A1, AAF and A2 were plotted in the oddball condition and control condition. (C) Comparision of calcium response to the first standard stimuli, pre-odd standard stimuli, and oddball stimuli across distinct auditory fields in omission condition. (D) The first derivatives of traces (dark: standard, yellow: pe_) from A1, AAF and A2 were plotted in the oddball condition and control condition.

To test for peak differences between stimulus types in the time-resolved responses to deviants, we conducted a two-way ANOVA (area*type) based on responses to the deviant elements in each pair. Here, we only analyzed peaks to the syllables with low transition probability. Specifically, we compared the derivative peak values within different time windows of the deviant stimulus ('da bi': the early and late peaks of the first syllable, with 3.93% transition probability pe bi→da; 'pe da': the early and late peaks of the second syllable, with 3.66% transition probability pe→da; 'bi pe': the early and late peaks of the first syllable, with 3.93% transition probability pe bi→bi) against the derivative peaks of the standard stimuli at the same time windows. The ANOVA was conducted separately for the oddball condition and control condition. Based on these analyses, we found a significant main effect of stimulus type in both conditions (oddball condition: $F_{(2,555)}$ =329.82, P<0.001, control condition: $F_{(2,555)}$ =143.84, P<0.001). The main effect of auditory areas was not significant in either condition (oddball condition: $F_{(2,555)}$ =0.35, P=0.705, control condition: $F_{(2,555)}$ =1.13, P=0.322). Importantly, however, the interaction effect between stimulus type and area was significant in the oddball condition ($F_{(4,555)}$ =2.66, P=0.032), but not in the control condition ($F_{(4,555)}$ =0.36, P=0.837).

## Discussion

In this study, we examined whether the auditory cortical activity in awake mice shows sensitivity to violations of sequences based on local stimulus probability only, or whether it is also sensitive to more complex violations based on stimulus order. We found that mice could encode both the local probabilities and the more global stimulus patterns and elicit mismatch signals to the substitution deviants (‘pe bi’→ ‘pe da’ or ‘da bi’) and transposition (‘pe bi’→‘bi pe’) deviants (Fig. 14A), but not to omission deviants (‘pe bi’→‘pe _’; Fig. 15d). Interestingly, the A2 area was found to elicit more pronounced MMRs to those deviants compared with A1 (Fig. 14A), which suggests a hierarchical gradient of prediction error signaling.

Consistent with previous studies in humans (Todorovic et al., 2011) and animals (Chen et al., 2015; Parras et al., 2017), we found a robust attenuation of responses to consecutive standards in all auditory fields (Fig. 13A). Similar effects have also been found in other modalities (Hamm et al., 2021; Hamm & Yuste, 2016; Homann et al., 2017; Rao & Ballard, 1999; Shipp et al., 2013). Although, based on the current experimental design, we cannot exclude the possibility that this repetition suppression effect was driven by passive adaptation (May & Tiitinen, 2010), since the temporal interval between stimulus pairs (~2350 ms) was within the recovery range (about <10s) of the synaptic depression (Ulanovsky et al., 2004), an increasing number of studies

(Tang et al., 2018; Todorovic & de Lange, 2012; Wacongne et al., 2012) suggest that repetition suppression is not only due to passive adaption, but rather it may reflect a gradual reduction of a prediction error signal.

In the analysis of responses to deviant stimuli, both substitution deviants ('pe da' and 'dabi') resulted in robust mismatch responses in all auditory fields when comparing deviants against the immediately preceding standards (Fig. 14A). These MMRs were larger in A2 than A1, suggesting a hierarchical gradient of deviance processing. Such a hierarchical organization is consistent with other studies (Parras et al., 2017) , which showed that higher-order regions elicit larger prediction error signals. Notably, a similar difference between auditory cortical regions (A2 > A1) was also identified when comparing neural response (AUC) to 'da bi' deviants (in which the first element of the pair was surprising) with the first standards in the sequence (Fig. 14C), i.e., controlling for adaptation effect. However, no such effect was found for 'pe da' deviants (in which the second element of the pair was surprising, see Fig. 14B). This result indicates that the serial position of the deviant element in the pair can modulate the prediction error, as the transitional probability between the surprising stimulus and the preceding stimulus is similar in both cases ('pe bi' followed by da bi, 3.93%, pe da, 3.66%). However, an alternative explanation might be that the neural habituation level is

different in these two conditions due to differences in time intervals between the deviant stimulus and the preceding stimulus (‘peda’: 145 ms gap between the expected ‘pe’ and the surprising ‘da’; ‘dabi’: 2350 ms gap between the pre-odd standards ‘bi’ and the surprising ‘da’). Interestingly, a recent electrophysiology study (Parras et al., 2021) suggests that PAF (the posterior auditory field) produces an even more robust prediction error response to single deviants than A2. Future studies on responses to more complex substitution deviants should expand the imaging range of cortical regions to include the PAF, to test whether the gradient of mismatch responses extends further beyond A2.

Interestingly, our results also show that both A2 and AAF elicit MMRs to the transposition deviants when compared against the immediately preceding standards. However, in this case we also identified a major contribution of repetition suppression to the mismatch signal, as no difference was found when comparing the deviants against the first standard. In contrast, no MMR was found in A1. The finding that non-A1 areas are sensitive to transposition deviants may be based on the local transitional probabilities (substitution deviant vs. the preceding standard: pebi→bi, 3.93%, bi→pe, 92.15%). However, it seems that the local probability of the stimulus is the most influential cue for the deviant processing in this study, as the MMRs were found to be much weaker for the transposition deviants than for substitution deviants, although the

transition probability between the surprising syllable and the preceding standard syllable was matched between deviant types equal (pebi→bi, 3.93%, pebi→da, 3.93%).

In contrast, our results based on responses to omission deviants (Fig. 15C) showed no robust mismatch response to the omitted stimulus. Responses to omissions might not be encoded in calcium activity (analyzed here) or spiking activity mainly from excitatory neurons in the superficial layers (to which wide-field imaging is most sensitive) (Waters, 2020), but rather in population activity or dendritic currents. Our devrivative results (see Fig. 15D) also suggested that there were no calcium events during the omitted stimulus (the second syllabic position), while calcium events at the second syllabic position were found in other types of stimuli and were modulated by the oddball types (see Fig. 15B). The omission stimulus is assumed to split the actively prediction signal from the prediction error (SanMiguel et al., 2013; Wacongne et al., 2012). Thus, while our results based on omission responses are inconsistent with the standard microcircuits proposed to implement predictive coding (Friston, 2005; Heilbron & Chait, 2018), more evidence is needed to elucidate the types of neurons and population mechanisms underlying omission signalling.

In summary, we show that mice can encode the deviants based on local stimulus

probabilities, but also based on transition probabilities related to the serial position of the deviants and the novelty of the deviant stimulus given preceding elements. Higher-order auditory regions (A2) show larger MMRs than primary regions (A1), which extends previous findings in traditional oddball paradigms based on local stimulus probabilities, showing an increase of mismatch signalling along the auditory processing pathway (Parras et al., 2017) . Our findings thus show that the rat auditory cortex is a suitable model also for deviant processing based on more complex transitional probabilities.

## Methods

### *Surgery for chronic window implantation*

Firstly, to prevent brain swelling during craniotomy, 0.1cc dexamethasone (2mg/ml, VetOne, USA) was administered subcutaneously at least two hours before surgery. The animals were initially anesthetized using 4% isoflurane (Fluriso, VetOne, USA) with a calibrated vaporizer (Matrix VIP 3000 Vaporizer, Midmark Corporation, USA). Isoflurane concentration was then lowered down to 1.5%–2% during surgery, and the depth of anesthesia was continuously monitored by toe pinch. Body temperature was maintained at 36.0 $^{O}$C during surgery. Hair remover face cream (Nair, Church and Dwight, USA) was applied on the surgical area (the top and left side of the animal's

head) to remove hair thoroughly. Povidone-iodine and 70% ethanol were applied to clean and sanitize the surgical area. Eye gel was applied to prevent eye drying. Then, a longitudinal curve scalp incision was made for skull exposure. Soft tissue under the scalp and periosteum was scraped by a scalpel. Then a circular cranial window was drilled over the left auditory cortex and is about 3.5mm in diameter. The dura matter was gently removed, and gelatin sponges were applied to stop bleeding. A sterile coverslip with three layers, which was made by the same method as (Liu et al., 2019) described, was implanted onto the brain surface over the left auditory cortex as a chronic recording window. Surgical silicone adhesive (kwik-sil, World Precision Instruments, USA) and quick adhesive cement (C&B Metabond, USA) were applied for window implantation. Lastly, a customized stainless-steel head plate was mounted on the top of the animal's head by using C&B Metabond to secure the head for chronic imaging. Antibiotic, 0.05cc Cefazolin (500 mg/2.2 ml, 1 g/vial, West-Ward Pharmaceuticals, USA), was injected subsequently. All experimental procedures were approved by the University of Maryland's Animal Care and Use Committee.

After the surgery was completed, the animals were kept individually and stayed warm under heating lights until regaining consciousness. Antibiotics were prepared by diluting 6 ml of Sulfamethoxazole and Trimethoprim Oral Suspension (200mg Sulfamethoxazole and 40mg Sulfamethoxazole per 5ml, Aurobindo Pharma, India) in

100 ml of water, and were administered to the mice for a week following the surgery. There were no imaging experiments during recovery (1-2 weeks).

***Experimental design and Auditory stimuli***

In the FRA mapping blocks, pure tones were generated with custom MATLAB scripts. Each tone lasted 2 s with linear ramps of 5 ms at the beginning and at the end of the tone. Nine tones with equal logarithmic spacing between 4 and 64 kHz were used at three attenuation levels: 0 dB, 20 dB, and 40 dB. Each stimulus was repeated 10 times. Stimuli were presented in a random order.

In the experimental blocks, artificial acoustic syllables pairs were used to form sequences containing standards and deviants. The syllables were selected from a database of CV syllables recorded by a male speaker (Ives et al., 2005) and were then analyzed and resynthesized by an open-source vocoder, STRAIGHT (Kawahara, 2006), in MATLAB R2018b (MathWorks Inc., Natick, USA). To generate experimental stimuli, we first manipulated and matched the stimulus onset and duration of all syllables (syllable duration = 251 ms), and then shifted the fundamental frequency and formant scalar of each CV syllable upward 1 octave to match the optimal rat hearing range (Kelly & Masterton, 1977). In the main experiment (the deviant condition), we

used ‘pebi’ as the standard stimulus (365 repeats, 85.88%), and ‘peda’ (15 repeats, 3.53%), ‘dabi’ (15 repeats, 3.53%), ‘bipe’(15 repeats, 3.53%), and ’pe_’ (15 repeats, 3.53%) as the deviant stimuli. In the control condition, all stimuli were presented with equal probability and repeated 20 times in a random order. The Stimulus-onset asynchrony (SOA) between pairs was 3 s and the gap between pair elements (offset to onset) was 145ms.

The amplitudes of all auditory stimuli were calibrated to 75dB SPL with a microphone (Bruel & Kjær 4944-A, Denmark). During sound presentation, sound waveform was loaded into RX6 multi-function processor (Tucker-Davis Technologies (TDT)) and attenuated to desired sound levels by PA5 attenuator (TDT). Then the signal was fed into the ED1 speaker driver (TDT), which drove an ES1 electrostatic speaker (TDT). The speaker was placed on the right-hand side of the animal, 10 cm away from the head, at an angle of 45 degrees relative to the mid-line.

***Widefield imaging and data preprocessing***

Mice (Thy1-Gcamp6s) were head-fixed by a customized head holder and placed on a rotatable platform beneath Ultima-IV microscope (Bruker Technologies). We used 470 nm LED light (M470L3, Thorlabs Inc.) to excite green fluorescence and acquired the

images (**10Hz,** 100ms exposure time, 400 * 400 pixels) with StreamPix 6.5 software (Norpix). Our imaging modality was also single photon, the calcium signals should mainly originate from the L2/3 (Waters, 2020; West et al., 2021). We down sampled the original image using the MATLAB built-in function 'imresize' from 400 pixels by 400 pixels to 100 pixels by 100 pixels. Next, we performed whitening of the image sequence and image segmentation by using exactly the same procedures as previously described (Liu et al., 2019), including a constrained autoencoder. Finally, we chose the first 50 components generated by the autoencoder. In the group-level analysis, we compared the AUC of traces and first derivative peaks across areas and stimuli types by using the two-way ANOVA. In addition, we compared the difference between deviants with standards across areas by using paired t-test. The false discovery rate (FDR < 0.05) was applied for correction for multiple comparisons.

## References


An, H., Auksztulewicz, R., Kang, H., & Schnupp, J. W. H. (2021, Jan). Cortical mapping of mismatch responses to independent acoustic features. *Hear Res, 399*, 107894. https://doi.org/10.1016/j.heares.2020.107894

Auksztulewicz, R., & Friston, K. (2015, Nov). Attentional Enhancement of Auditory Mismatch Responses: a DCM/MEG Study. *Cereb Cortex, 25*(11), 4273-4283. https://doi.org/10.1093/cercor/bhu323

Auksztulewicz, R., & Friston, K. (2016, Jul). Repetition suppression and its contextual determinants in predictive coding. *Cortex, 80*, 125-140. https://doi.org/10.1016/j.cortex.2015.11.024

Cappotto, D., Auksztulewicz, R., Kang, H., Li, K., Melloni, L., & Schnupp, J. (2021). Simultaneous mnemonic and predictive representations in the auditory cortex. *bioRxiv*.

Carbajal, G. V., & Malmierca, M. S. (2018). The neuronal basis of predictive coding along the auditory pathway: from the subcortical roots to cortical deviance detection. *Trends in hearing, 22*, 2331216518784822.

Chao, Z. C., Takaura, K., Wang, L., Fujii, N., & Dehaene, S. (2018, Dec 5). Large-Scale Cortical Networks for Hierarchical Prediction and Prediction Error in the Primate Brain. *Neuron, 100*(5), 1252-1266 e1253. https://doi.org/10.1016/j.neuron.2018.10.004

Chen, I. W., Helmchen, F., & Lutcke, H. (2015, Sep 9). Specific Early and Late Oddball-Evoked Responses in Excitatory and Inhibitory Neurons of Mouse Auditory Cortex. *J Neurosci, 35*(36), 12560-12573. https://doi.org/10.1523/JNEUROSCI.2240-15.2015

Chennu, S., Noreika, V., Gueorguiev, D., Shtyrov, Y., Bekinschtein, T. A., & Henson, R. (2016). Silent expectations: dynamic causal modeling of cortical prediction and attention to sounds that weren't. *Journal of Neuroscience, 36*(32), 8305-8316.

Chouiter, L., Tzovara, A., Dieguez, S., Annoni, J.-M., Magezi, D., De Lucia, M., &

Spierer, L. (2015). Experience-based auditory predictions modulate brain activity to silence as do real sounds. *Journal of cognitive neuroscience, 27*(10), 1968-1980.

Cowan, N., Winkler, I., Teder, W., & Naatanen, R. (1993, Jul). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *J Exp Psychol Learn Mem Cogn, 19*(4), 909-921. https://doi.org/10.1037//0278-7393.19.4.909

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015, Oct 7). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron, 88*(1), 2-19. https://doi.org/10.1016/j.neuron.2015.09.019

Friston, K. (2005, Apr 29). A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci, 360*(1456), 815-836. https://doi.org/10.1098/rstb.2005.1622

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clinical neurophysiology, 120*(3), 453-463.

Hamm, J. P., Shymkiv, Y., Han, S., Yang, W., & Yuste, R. (2021, Apr 6). Cortical ensembles selective for context. *Proc Natl Acad Sci U S A, 118*(14). https://doi.org/10.1073/pnas.2026179118

Hamm, J. P., & Yuste, R. (2016). Somatostatin interneurons control a key component of mismatch negativity in mouse visual cortex. *Cell reports, 16*(3), 597-604.

Heilbron, M., & Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex? *Neuroscience, 389*, 54-73.

Herrmann, B., Henry, M. J., Fromboluti, E. K., McAuley, J. D., & Obleser, J. (2015). Statistical context shapes stimulus-specific adaptation in human auditory cortex. *Journal of neurophysiology, 113*(7), 2582-2591.

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

Ives, D. T., Smith, D. R., & Patterson, R. D. (2005, Dec). Discrimination of speaker size from syllable phrases. *J Acoust Soc Am, 118*(6), 3816-3822. https://doi.org/10.1121/1.2118427

Kawahara, H. (2006). STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. *Acoustical science and technology, 27*(6), 349-353.

Kelly, J. B., & Masterton, B. (1977). Auditory sensitivity of the albino rat. *Journal of comparative and physiological psychology, 91*(4), 930.

Liu, J., Whiteway, M. R., Sheikhattar, A., Butts, D. A., Babadi, B., & Kanold, P. O. (2019). Parallel processing of sound dynamics across mouse auditory cortex via spatially patterned thalamic inputs and distinct areal intracortical circuits. *Cell reports, 27*(3), 872-885. e877.

Luo, D., Li, K., An, H., Schnupp, J. W., & Auksztulewicz, R. (2021). Learning boosts the decoding of sound sequences in rat auditory cortex. *Current Research in Neurobiology, 2*, 100019.

May, P. J., & Tiitinen, H. (2010, Jan 1). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology, 47*(1), 66-122. https://doi.org/10.1111/j.1469-8986.2009.00856.x

Naatanen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007, Dec). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol, 118*(12), 2544-2590. https://doi.org/10.1016/j.clinph.2007.04.026

Parras, G. G., Casado-Roman, L., Schroger, E., & Malmierca, M. S. (2021, Nov 15). The posterior auditory field is the chief generator of prediction error signals in the auditory cortex. *Neuroimage, 242*, 118446. https://doi.org/10.1016/j.neuroimage.2021.118446

Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdes-Baizabal, C., Escera, C., & Malmierca, M. S. (2017, Dec 15). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nat Commun, 8*(1), 2148.

https://doi.org/10.1038/s41467-017-02038-6

Perez-Gonzalez, D., Malmierca, M. S., & Covey, E. (2005, Dec). Novelty detector neurons in the mammalian auditory midbrain. *Eur J Neurosci, 22*(11), 2879-2885. https://doi.org/10.1111/j.1460-9568.2005.04472.x

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience, 2*(1), 79-87.

Romero, S., Hight, A. E., Clayton, K. K., Resnik, J., Williamson, R. S., Hancock, K. E., & Polley, D. B. (2020, Mar 14). Cellular and Widefield Imaging of Sound Frequency Organization in Primary and Higher Order Fields of the Mouse Auditory Cortex. *Cereb Cortex, 30*(3), 1603-1622. https://doi.org/10.1093/cercor/bhz190

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schröger, E. (2013). Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. *Journal of Neuroscience, 33*(20), 8633-8639.

Shipp, S., Adams, R. A., & Friston, K. J. (2013). Reflections on agranular architecture: predictive coding in the motor cortex. *Trends in neurosciences, 36*(12), 706-716.

Tang, M. F., Smout, C. A., Arabzadeh, E., & Mattingley, J. B. (2018). Prediction error and repetition suppression have distinct effects on neural representations of visual information. *Elife, 7*, e33123.

Todorovic, A., & de Lange, F. P. (2012, Sep 26). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J Neurosci, 32*(39), 13389-13395. https://doi.org/10.1523/JNEUROSCI.2227-12.2012

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *Journal of Neuroscience, 31*(25), 9118-9123.

Uhrig, L., Dehaene, S., & Jarraya, B. (2014, Jan 22). A hierarchy of responses to auditory regularities in the macaque brain. *J Neurosci, 34*(4), 1127-1132. https://doi.org/10.1523/JNEUROSCI.3165-13.2014

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004, Nov 17). Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci, 24*(46), 10440-10453. https://doi.org/10.1523/JNEUROSCI.1905-04.2004

Ulanovsky, N., Las, L., & Nelken, I. (2003, Apr). Processing of low-probability sounds by cortical neurons. *Nat Neurosci, 6*(4), 391-398. https://doi.org/10.1038/nn1032

Wacongne, C., Changeux, J. P., & Dehaene, S. (2012, Mar 14). A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci, 32*(11), 3665-3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011, Dec 20). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc Natl Acad Sci U S A, 108*(51), 20754-20759. https://doi.org/10.1073/pnas.1117807108

Wang, L., Uhrig, L., Jarraya, B., & Dehaene, S. (2015). Representation of numerical and sequential patterns in macaque and human brains. *Current Biology, 25*(15), 1966-1974.

Waters, J. (2020, Nov 6). Sources of widefield fluorescence from the brain. *Elife, 9*. https://doi.org/10.7554/eLife.59841

West, S. L., Aronson, J. D., Popa, L. S., Feller, K. D., Carter, R. E., Chiesl, W. M., Gerhart, M. L., Shekhar, A. C., Ghanbari, L., Kodandaramaiah, S. B., & Ebner, T. J. (2021, Oct 23). Wide-Field Calcium Imaging of Dynamic Cortical Networks during Locomotion. *Cereb Cortex*. https://doi.org/10.1093/cercor/bhab373

Yang, T., Hamalainen, J. A., Lohvansuu, K., Lipponen, A., Penttonen, M., & Astikainen, P. (2021, Jan). Deviance detection in sound frequency in simple and complex sounds in urethane-anesthetized rats. *Hear Res, 399*, 107814. https://doi.org/10.1016/j.heares.2019.107814

Yaron, A., Hershenhoren, I., & Nelken, I. (2012, Nov 8). Sensitivity to complex statistical regularities in rat auditory cortex. *Neuron, 76*(3), 603-615.

https://doi.org/10.1016/j.neuron.2012.08.025

# Chapter 4. Congruency of "What" and "When" Predictions in the Auditory System


## Abstract

The brain's ability to form predictions based on statistical regularities in ongoing stimulus teams is an integral part of forming adaptive behaviors. Such events form regularities both in terms of the characteristics of the stimuli itself (e.g. "what" that event is) and in the predictability of onset for when events occur within a sequence (e.g. "when" the predicted event will take place). In real-world stimulus streams, regularities also occur at different hierarchies of complexity and meaning - e.g. predictions of individual notes vs. melodic contour, in the case of music perception, or vowel vs. sentence syntax predictions in natural language processing. However, the underlying mechanisms between "what" and "when" predictive processing have been traditionally investigated using non-complementary experimental paradigms, preventing their full disambiguation. To address this, the present study employs "what" and "when" prediction violations at different hierarchies within the same stimulus stream during an oddball task, while recording neural activity in human participants via electroencephalogram. Our results revealed a congruent effect in mismatch response amplitude between hierarchically structured "what" and "when" violations, supported

by the presence of decreased left parietal activity.

## Introduction

The ability to predict future events based on incoming stimulus streams is an integral aspect of sensory processing. Real-world sensory events contain multiple features which can be used to form such predictions, and humans can predict information in a sequence based on statistical regularities and chunking (Dehaene et al., 2015). A recent study (Ding,et al. 2016) has similarly suggested that cortical activity can entrain to hierarchical structures in linguistic sequences pursuant to levels of chunking. Typically, studies of predictive coding in the auditory system manipulate features that are content-based or time-based in predictability. Such "what" and "when" predictions are present in virtually all auditory stimulus streams, and their manipulation has been the foundation for numerous classical and contemporary studies (Denham and Winkler, 2020). In the case of oddball paradigms often employed in predictive coding research, a deviant stimulus token is presented in place of an otherwise predictable token within a repeated stream, resulting in the classical MMR and its modern interpretation as an error correction signal within predictive coding frameworks (Garrido et al., 2009). However, to make a prediction about what token will occur next in a sequence

necessitates a mechanism for predicting *when* that token will occur within a regular stream. As opposed to "what" predictions, which often rely on MMR-based explanations, "when" predictions are typically explained by the concept of neural entrainment - phase alignment of neural activity to an external temporal structure (Auksztulewicz et al., 2019; Haegens and Zion Golumbic, 2018; Schroeder and Lakatos, 2009).

Recent cross-modal studies have investigated these disparate mechanisms through independent manipulation of timing and content predictability, observing partly dissociable neural correlates and putative underlying mechanisms. In an audio-visual study (Auksztulewicz et al., 2018), the predictability of auditory targets was modulated by the preceding visual cues, and it was observed that content and timing predictability evoked activity in dissociable regions and time windows, while also interactively modulating activity in the temporal gyrus during early latencies. Timing and content modulations have been observed in audio-motor linguistic studies, with overlapping "what" and "when" predictive activity observed only in higher level regions (Emmendorfer et al., 2020). Similar phenomena have been observed in lemniscal and non-lemniscal oscillatory activity associated with the temporal and syntactic content of musical sequences (Musacchia et al., 2014). In this context, neural entrainment of non-

lemniscal regions representing stimulus rhythms can be seen to modulate lemniscal regions representing stimulus contents, including MMR processing along the ascending pathways. However, it is unknown if these interactions are specific to the hierarchical levels present in complex naturalistic stimuli such as speech or music (Hasson et al., 2015). MMR amplitude has been shown to increase as a function of deviance, with deviants straying further from the standard eliciting higher MMR amplitudes (Jalewa et al., 2020). This phenomenon corresponds not only to stimulus characteristics such as the frequency of a deviant relative to the expected standard, but also to stimulus onset asynchrony, with deviants occuring at the expected temporal onset eliciting higher amplitudes than those at jittered onsets (Auksztulewicz et al., 2019, 2018; Jalewa et al., 2020; Musacchia et al., 2014). In the case of naturalistic music, lower-level contents are present in the form of single notes within a sequence and higher-level contents present in the resulting melody of a sequence, each occurring at their respective time scales. Based on current theories it is unclear if entrainment modulates processing of stimulus contents in a hierarchically specific way - e.g. if entrainment amplifies the processing of stimuli presented at a preferred time window, or only to inputs presented at a relevant time scale.

Here, we disambiguate neural correlates of “what” and “when” predictions by

independently manipulating temporal and syntactical content at two hierarchical levels. Human EEG signals were recorded during an auditory oddball task in order to gauge the effect of temporal predictive mechanisms at hierarchically higher and lower temporal scales, and their modulatory effect on higher and low-level content-based predictive hierarchies. Musical sequences (ascending or descending musical scales) were chosen as stimulus sets, to reduce the influence of speech-specific processing on neural activity (e.g., modulation by language comprehension, speech-specific semantic and syntactic processing, etc.) and provide a better comparison to similar work in animal models. Temporal predictability was manipulated at slower (2 Hz) and faster (4 Hz) time scales, while acoustic deviants were introduced at lower (e.g. single tones) and higher (e.g. tone pairs) hierarchical levels. In the analysis, we focused on interactions between "what" and "when" predictions, specifically testing whether MMRs are modulated by temporal predictability in a hierarchically specific way. To explain the effects observed at the scalp level, we used source reconstruction, which allowed us to infer the putative mechanisms of interactions between "what" and "when" predictions.

## Methods

### *Participant sample*

Participants (N=20, median age 21, range 19-25), 10 females, 10 males; 19 right-handed, 1 left-handed) volunteered to take part in the study and gave written consent. The work was conducted in accordance with protocols approved by the Human Subjects Ethics Sub-Committee of the City University of Hong Kong. All participants self-reported normal hearing and no current or past neurological or psychiatric disorders.

***Stimulus design and behavioral paradigm***

Participants were exposed to auditory sequences containing independent manipulations of "what" and "when" predictability. Stimuli were presented in sequences (trials) of 7 ascending or descending scales, each composed of 8 musical notes with fundamental frequencies increasing or decreasing in 8th octave steps, so that one scale covered one octave. Each sequence thus comprised a total of 56 notes(Fig. 11A, 11B). Each participant heard a total of 240 sequences (trials). Within a trial, all scales were either ascending or descending, and the ascending or descending trials were presented in a random order. The fundamental frequency of the initial note of each scale was randomly drawn from a range 300-600 Hz. Each tone was generated by resynthesizing a virtual harp note F4 (played on virtualpiano.net), to match a fixed 166 ms duration and the desired fundamental frequency. The required manipulations of the original note were implemented in an open-source vocoder, STRAIGHT (Kawahara, 2006) for Matlab

2018b (MathWorks; RRID: SCR_001622). Tones were perceptually grouped into pairs by manipulating the intensity ratio of odd/even tones, with the even (2nd, 4th, 6th and 8th) tones within a scale presented 10 dB quieter relative to the odd-position tones.

The temporal predictability of the sequences was manipulated according to three "predictability conditions", which were administered in 12 blocks of 20 trials (4 blocks per condition). The blocks were presented in a pseudo-random order, allowing no immediate repetitions of the same condition. In the "fully predictable" (isochronous) condition, tones were presented with a fixed SOA of 247 ms, resulting in all notes having predictable timing at both the slow time scale (tone pairs) and the fast time scale (single tones). In the "predictable slow, unpredictable fast" condition, the slow time scale was predictable (corresponding to a fixed pair onset asynchrony, i.e., a fixed 494 ms interval between the onsets of the odd, pair-initial tones) but the fast time scale was unpredictable (corresponding to a random onset of the even, pair-final tones, relative to the pair-initial tones). In this condition, the exact SOA of the pair-final tones was set by randomly drawing one value from the following 4 SOAs, relative to the standard 247 ms SOA: 33.3% shorter; 16.6% shorter; 16.6% longer; 33.3% longer. Finally, in the "predictable fast, unpredictable slow" condition, the fast time scale was predictable (corresponding to a fixed 247 ms SOA of the pair-final tones, relative to the pair-initial tones) but the slow time scale was unpredictable (corresponding to a random onset of

the odd, pair-initial tones, relative to the expected 494 ms interval). In this condition, the exact SOA of the pair-initial tones was set by randomly drawing one value from the same 4 SOAs as above, and shifting the onset of the pair-initial tone by this value, relative to the expected 494 ms interval relative to the previous pair onset. A fixed inter-trial interval of 1 second was employed between the offset of the last tone of a 56-tone sequence and the onset of the first tone in the next sequence.

Content predictability was manipulated by altering the fundamental frequency of a subset of tones within the scales, such that trials could contain a "fast" deviant (i.e., a single deviant tone) or a "slow" deviant (i.e., a deviant tone pair). The "fast" deviants were introduced by replacing the final tone of a scale with an outlier frequency (i.e., a tone whose fundamental frequency was 20% lower/higher than the range of the entire scale). The "slow" deviants were introduced by replacing the penultimate tone of a scale (i.e., the initial tone of the final pair) in the same manner. In each trial, the first two scales were left unaltered (to facilitate the extraction of statistical regularities in the sequences), and two deviant tones were randomly placed within the subsequent 5 scales. Additionally, in 50% of the trials, a scale containing an immediate tone repetition was included in the last 5 scales. In subsequent EEG analysis, neural responses evoked by "fast" and "slow" deviants were compared with neural responses evoked by the

respective standard tones, designated as the final (“fast”) and penultimate (“slow”) tones in two unaltered scales out of the final 5.

In total, 64.3% of the scales were left unaltered, 14.3% contained a “fast” deviant, 14.3% contained a “slow” deviant, and 7.1% contained a tone repetition. The global deviant probability equaled 3.57% of all tones, amounting to 80 deviant tones per deviant type (“fast”, “slow”) per temporal condition (“fully predictable”, “predictable fast”, “predictable slow”). To ensure that the EEG analysis is not confounded by differences in baseline duration between temporal conditions (e.g., “fast” deviants preceded by shorter/longer SOAs in the “predictable slow” condition than in the other two conditions), the SOAs preceding all deviant tones and designated standard tones were replaced by a fixed 247 ms SOA. Therefore, the temporal predictability manipulation was purely contextual, and did not affect the exact timing of either deviants or standards.

Auditory sequences were generated using the freely-available Psychtoolbox for MATLAB and delivered to participants fitted with Brainwavz B100 earphones via a TDT RZ6 multiprocessor at a playback sampling rate of 24414 Hz. Participants were seated in a sound-attenuated EEG booth. Visual stimuli (fixation cross) and instructions

were presented on a 24-inch computer monitor and delivered using the Psychophysics Toolbox for MATLAB. Participants were asked to minimize movements and eye blinks and instructed to perform a tone repetition detection task, by pressing a button as soon as possible upon hearing an immediate tone repetition. Prior to experimental blocks, participants were exposed to a training session consisting of "fully predictable" sequences containing a tone repetition, to familiarize themselves with the task and stimuli. Participants performed training trials until they could detect tone repetition in 3 consecutive trials with reaction times shorter than 2 seconds. Then, during the actual experiment, participants received feedback on their mean accuracy and reaction time after each block of 20 trials. The data segments (scales) containing tone repetition were subsequently rejected from EEG analysis.

***Behavioral analysis***

Analysis was performed on the accuracy and reaction time data corresponding to participant responses during the repetition detection task. Reaction times longer than 2 seconds were excluded from analysis, and mean reaction times were log-transformed to approximate a normal distribution. Accuracy and mean reaction times were entered into separate repeated-measures ANOVAs with a within-subjects factor Time (fully predictable, predictable fast, predictable slow). Post-hoc comparisons were

implemented using paired t-tests.

**Figure 16**

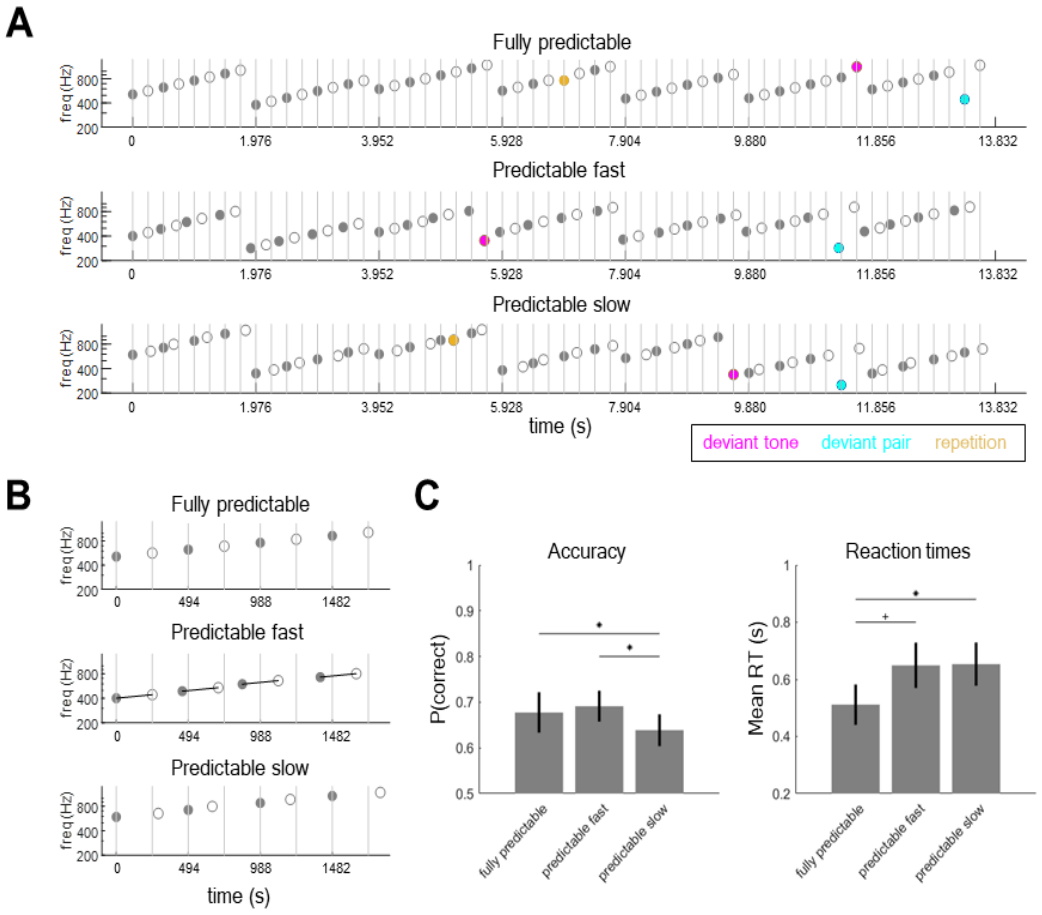


**Figure 16. Experimental paradigm and behavioral results. (A)** Participants listened to sequences of ascending (pictured) or descending scales of acoustic tones. Sequences were composed of tone pairs, where odd tones (grey circles) were louder than even tones (white circles). Participants performed a tone repetition detection task (orange circles: behavioral targets). Additionally, sequences could include deviant tones (red circles), in which one of the pair-final tones had an outlier frequency, and deviant pairs (blue circles), in which one of the pair-initial tones had an outlier frequency. **(B)** Sequences were blocked into three temporal conditions: a "fully predictable" condition (upper panel), in which ISI between tones was fixed at 0.25 s; a "predictable fast" condition (middle panel), in which the ISI between odd and even tones within pairs was

fixed at 0.25 s but the ISI between odd tones (pair-initial tones) was jittered; and a "predictable slow" condition (lower panel), in which ISI between odd tones (pair-initial tones) was fixed at 0.5 s but the ISI between odd and even tones within pairs was jittered. **(C)** Behavioral results. Left panel: accuracy, right panel: reaction times. Error bars denote SEM across participants. Asterisks denote $p < 0.05$, plus symbol denotes a trend towards significance.

***Neural data acquisition and pre-processing***

EEG signals were collected using a 64-channel ANT Neuro EEGo Sports amplifier at a sampling rate of 1024 Hz. The signals were grounded at the nasion and referenced to the Cpz electrode. The recorded data were pre-processed using the SPM12 Toolbox (Wellcome Trust Centre for Neuroimaging, University College London; RRID: SCR_007037) for MATLAB. Continuous data were high-pass filtered at 0.1 Hz and notch filtered 48 Hz and 52 Hz before being down-sampled to 300 Hz and subsequently low-pass filtered at 90 Hz. All filters were 5th order zero-phase Butterworth. Eyeblink artifacts were removed using channel Fpz and subtracting their two spatiotemporal principal components from all EEG channels (Ille et al., 2002). Cleaned signals were re-referenced to the average of all channels. The pre-processed data were analyzed separately in the frequency domain (phase coherence analysis) and in the time domain (event-related potentials; ERPs).

***Phase coherence analysis***

To test whether tone sequences are associated with dissociable spectral peaks in the neural responses at the single-tone rate (4.048 Hz) and at the tone-pair rate (2.024 Hz), we analyzed the data in the frequency domain. Continuous data were segmented into epochs ranging from the onset to the offset of each trial (tone sequence). For each participant, channel, and sequence, we calculated the Fourier spectrum of EEG signals measured during that sequence. We then calculated the ITPC, separately for each temporal condition ("fully predictable", "predictable fast", "predictable slow") according to the following equation (Ding and Simon, 2013):

$$ITPC_f = \left(\left[\Sigma^N cos\phi_f\right]^2 + \left[\Sigma^N sin\phi_f\right]^2\right)/N,$$

where $\varphi_f$ corresponds to the Fourier phase at a given frequency *f,* and *N* corresponds to the number of sequences (80 per condition).

In the initial analysis, ITPC estimates were averaged across EEG channels. To test for the presence of statistically significant spectral peaks, ITPC values at the single-tone rate (4.048 Hz) and tone-pair rate (2.024 Hz) were compared against the mean of ITPC values at their respective neighboring frequencies (single-tone rate: 3.974 and 4.124 Hz; tone-pair rate: 1.949 and 2.099 Hz) using paired t-tests.

Furthermore, to test whether tone-rate and pair-rate spectral peaks observed at single EEG channels show modulations due to temporal predictability, spatial topography maps of single-channel ITPC estimates were converted to 2D images, smoothed with a 5 x 5 mm full-width-at-half-maximum (FWHM) Gaussian kernel, and entered into repeated-measures ANOVAs (separately for tone-rate and pair-rate estimates) with a within-subjects factor Time (fully predictable, predictable fast, predictable slow), implemented in SPM12 as a general linear model (GLM). Statistical parametric maps were thresholded at $p < 0.001$ and corrected for multiple comparisons over space at a cluster-level $pFWE < 0.05$ under random field theory assumptions (Kilner et al., 2005).

Finally, to test whether spectral signatures of temporal predictability are modulated by experience with stimuli, we split the data into two halves (two consecutive bins of 40 trials), separately for each condition. Tone-rate and pair-rate ITPC estimates were averaged across EEG channels and compared separately for each of the two halves using repeated-measures ANOVAs with a within-subjects factor Time (fully predictable, predictable fast, predictable slow).

***Event-related potentials***

For the time-domain analysis, data were segmented into epochs ranging from -50 ms before to 247 ms after deviant/standard tone onset, baseline-corrected from -25 ms to

25 ms to prevent epoch contamination due to the temporally structured presentation (Fitzgerald et al., 2021), and denoised using the "Dynamic Separation of Sources" (DSS) algorithm (de Cheveigné and Simon, 2008). Condition-specific ERPs (corresponding to "fast"/"slow" deviants and the respective standards, presented in each of the three temporal conditions) were calculated using robust averaging across trials, as implemented in the SPM12 toolbox, and low-pass filtered at 48 Hz (5th order zero-phase Butterworth). The resulting ERPs were analyzed univariately to gauge the effects of "what" and "when" predictions on evoked responses. ERP data were converted to 3D images (2D: spatial topography; 1D: time), and the resulting images were spatially smoothed using a 5 x 5 mm FWHM Gaussian kernel. The smoothed images were entered into a GLM implementing a 3 x 3 repeated-measures ANOVA with a within-subject factors Contents (standard, deviant tone, deviant pair) and Time (fully predictable, predictable fast, predictable slow). Beyond testing for the two main effects and a general 3 x 3 interaction, we also designed a planned contrast quantifying the congruency effect (i.e., whether "when" predictions specifically modulate the amplitude of mismatch signals evoked by deviants presented at a time scale congruent with "when" predictions, i.e., deviant tones in the "predictable fast" condition and deviant pairs in the "predictable slow" conditions). To this end, we tested for a 2 x 2 interaction between Contents (deviant tone, deviant pair) and Time (predictable fast,

predictable slow). Statistical parametric maps were thresholded at $p < 0.001$ and corrected for multiple comparisons over space and time at a cluster-level $pFWE < 0.05$ under random field theory assumptions (Kilner et al., 2005).

***Brain-behavior correlations***

To test whether the effects of "what" and/or "when" predictions correlate with each other, as well as with behavioral benefits of "when" predictions in the repetition detection task, we performed a correlation analysis across participants. Given that we observed a significant congruency effect (i.e., a 2 x 2 interaction between "what" and "when" predictions; see Results and Fig. 18C), as well as effects of the time scale of "when" predictions (predictable fast vs. predictable slow) on ITPC (Fig. 16D) and behavioral accuracy (Fig. 16C), we included these measures in the correlation analysis. Furthermore, since we hypothesized that performance in the repetition detection task might be related to overall deviance detection, we also included a measure of "what" predictions (i.e., the overall difference between deviant- and standard-evoked ERPs; Fig. 18AB). Thus, for each participant, we calculated a single behavioral index (the difference between accuracy scores obtained in the predictable fast vs. predictable slow condition) and three neural indices: (1) the difference between deviant-evoked ERP amplitudes measured in the temporally congruent (deviant tones presented in the predictable fast condition; deviant pairs presented in the predictable slow condition)

and incongruent (deviant pairs presented in the predictable fast condition; deviant tones presented in the predictable slow condition) conditions, averaged across electrodes in the significant cluster; (2) the difference between the pair-rate ITPC values obtained for predictable fast vs. predictable slow conditions in the second half of the experiment (see Results); and (3) the difference between the absolute deviant-evoked and standard-evoked ERP amplitudes (averaged across significant channels and temporal conditions). We then fitted a linear regression model with three predictors (the three neural indices) regressed against the accuracy effect, and identified outlier participants using a threshold of Cook's distance exceeding 5 times the mean. Correlations between all measures were quantified using Pearson's *r* and corrected for multiple comparisons using Bonferroni correction.

***Source reconstruction***

Source reconstruction was performed under group constraints (Litvak and Friston, 2008) using empirical Bayesian beamformer (Belardinelli et al., 2012; Little et al., 2018; Wipf and Nagarajan, 2009) based on the entire post-stimulus time window (0-247 ms). Since in the ERP analysis (see Results) we identified two principal findings - namely a difference between ERPs evoked by deviants and standards, and an interaction between deviant type and temporal condition - we focused on comparing source estimates corresponding to these effects. In the analysis of the difference between deviants and

standards, source estimates were extracted for the 173-223 ms time window, converted into 3D images consisting of 3 spatial dimensions and smoothed with a 10 x 10 x 10 mm FWHM Gaussian kernel. Smoothed images were then entered into a GLM implementing a 3 x 3 repeated-measures ANOVA with within-subjects factors of Content (standard, deviant tone, deviant pair) and Time (fully predictable, predictable fast, predictable slow). In the analysis of the interaction between deviant type and temporal condition, source estimates were extracted for the 130-180 ms and processed as above. Smoothed images were then entered into a GLM implementing a 2 x 2 repeated-measures ANOVA with within-subjects factors of Content (deviant tone, deviant pair) and Time (predictable fast, predictable slow). Statistical parametric maps were thresholded and corrected for multiple comparisons over space at a cluster-level $pFWE < 0.05$ under random field theory assumptions (Kilner et al., 2005). Source labels were assigned using the Neuromorphometrics probabilistic atlas, as implemented in SPM12.

## Results

### *Behavioral results*

Performance across all trials revealed significant differences in accuracy across conditions (main effect of Time: $F_{2,38} = 7.3530$, $p = 0.002$), corresponding to

significantly lower accuracy in the “predictable slow” condition (mean ± SEM: 63.88% ± 3.65%) than in the “fully predictable” (mean ± SEM: 67.75% ± 4.64%; t19 = -2.5272, p = 0.0205) and “predictable fast” conditions (mean ± SEM: 69.12% ± 3.55%; t19 = -5.984, $p < 0.001$) (Fig. 16C). Reaction times also significantly differed across conditions ($F_{2,38} = 3.5543$, p = 0.0385), with post-hoc analysis revealing that reaction times were significantly faster in the “fully predictable” condition (mean ± SEM: 511 ± 74 ms) than in the “predictable slow” condition (mean ± SEM: 653 ± 79 ms; t19 = 2.4089, p = 0.0263). The difference between the “fully predictable” condition and the “predictable fast” condition (mean ± SEM: 649 ± 83 ms) trended towards significance (t19 = 2.0132, p = 0.0585). No significant difference was observed between the “fully predictable” condition and the “predictable fast” condition (p = 0.9013).

***Phase coherence analysis***

In the EEG spectrum of inter-trial phase coherence (ITPC; averaged across conditions and channels), both tone-rate peak (4.048 Hz) and pair-rate peak (2.024 Hz) were observed, relative to neighboring frequency points (tone-rate: $t_{19} = 6.8489$, $p < 0.001$; pair-rate: $t_{19} = 3.6274$, $p = 0.0018$). The ITPC peak estimates differed between experimental conditions, reflecting differences in the stimulus spectrum. Specifically, the tone-rate ITPC estimates were higher in the “fully predictable” and “predictable slow” conditions than in the “predictable fast” conditions, and this effect was observed

at most of the EEG channels ($F_{max}$ = 46.30, $Z_{max}$ = 6.43, $p_{FWE}$ < 0.001; pairwise comparisons: “fully predictable” vs. “predictable fast”, $T_{max}$ = 8.02, $Z_{max}$ = 6.10, $p_{FWE}$ < 0.001; “predictable slow” vs. “predictable fast”, $T_{max}$ = 9.62, $Z_{max}$ = 6.81, $p_{FWE}$ < 0.001; “fully predictable” vs. “predictable slow”, all $p_{FWE}$ > 0.05). Conversely, the pair-rate ITPC estimates were higher in the “predictable slow” condition than in the other two conditions, and this effect was observed over right lateral channels ($F_{max}$ = 7.45, $Z_{max}$ = 2.90, $p_{FWE}$ = 0.031; pairwise comparisons: “predictable slow” vs. “fully predictable”, $T_{max}$ = 3.81, $Z_{max}$ = 3.48, $p_{FWE}$ = 0.004; “predictable slow” vs. “predictable fast”, $T_{max}$ = 3.83, $Z_{max}$ = 3.50, $p_{FWE}$ = 0.001; “fully predictable” vs. “predictable fast”, all $p_{FWE}$ > 0.05). Interestingly, the pair-rate differences between conditions built up during the experiment: they were absent during the first half of the experiment ($F_{2,59}$ = 1.0433, p = 0.3622), and were only observed during the second half of the experiment ($F_{2,59}$ = 3.8798, p = 0.0293). This was not the case for the tone-rate differences between conditions, which were stable during the experiment (first half: $F_{2,59}$ = 26.1701, p < 0.001; second half: $F_{2,59}$ = 26.9480, p < 0.001).

**Figure 17**

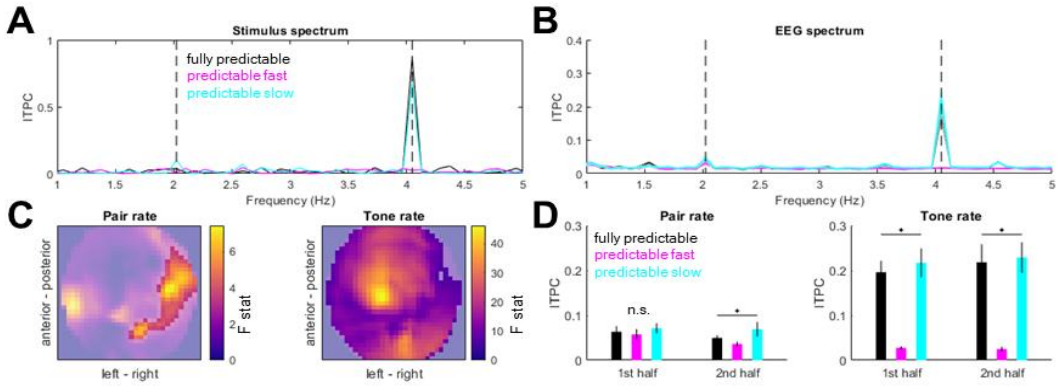


**Figure 17. Spectral signatures of temporal predictability.** (A) Inter-trial phase coherence (ITPC) in the stimulus spectrum. Black: "fully predictable", cyan: "predictable fast", magenta: "predictable slow". Pair-rate (2.024 Hz) and tone-rate (4.048 Hz) peaks are indicated by dashed vertical lines. (B) ITPC based on EEG activity (averaged across channels). Legend as above. Shaded areas indicate SEM across participants. (C) EEG topography maps of main effects of Condition (fully predictable vs. predictable fast vs. predictable slow) on the pair-rate peak ITPC (left panel) and tone-rate peak ITPC (right panel). Statistical F values are represented on the color scale. Unmasked area corresponds to significant clusters ($pFWE < 0.05$). (D) Pair-rate (left panel) and tone-rate (right panel) peak ITPC values plotted separately for the 1st half and 2nd half of the trials. Error bars denote SEM across participants.

***Event-related potentials***

To test for effects of "what" and "when" predictions on ERP amplitudes, we analyzed the data in the time domain. ERP amplitudes differed significantly between deviant and standard tones, pooled over temporal conditions (Fig. 18A; posterior cluster: 173-223 ms, $F_{max} = 53.94$, $Z_{max} = 6.68$, $p_{FWE} < 0.001$; anterior cluster: 177-220 ms; $F_{max} = 37.57$;

$Z_{max}$ = 5.67; $p_{FWE}$ < 0.001). When analysing specific deviant types ("fast" and "slow" deviants vs. their respective standards), significant differences between deviants and standards were observed in both cases ("fast" deviants vs. standards: posterior cluster, 173-223 ms, $F_{max}$ = 41.50, $Z_{max}$ = 5.94, $p_{FWE}$ < 0.001; anterior cluster, 177-227 ms; $F_{max}$ = 35.56; $Z_{max}$ = 5.52; $p_{FWE}$ < 0.001; "slow" deviants vs. standards: posterior cluster, 170-220 ms, $F_{max}$ = 45.63, $Z_{max}$ = 6.20, $p_{FWE}$ < 0.001; anterior cluster, 177-213 ms; $F_{max}$ = 30.17; $Z_{max}$ = 5.11; $p_{FWE}$ < 0.001). No significant differences were observed between the two deviant types, pooling over temporal conditions ($p_{FWE}$ > 0.05). Thus, the main effect of "what" predictions differentiated between deviants and standards, but not between deviant types.

**Figure 18**

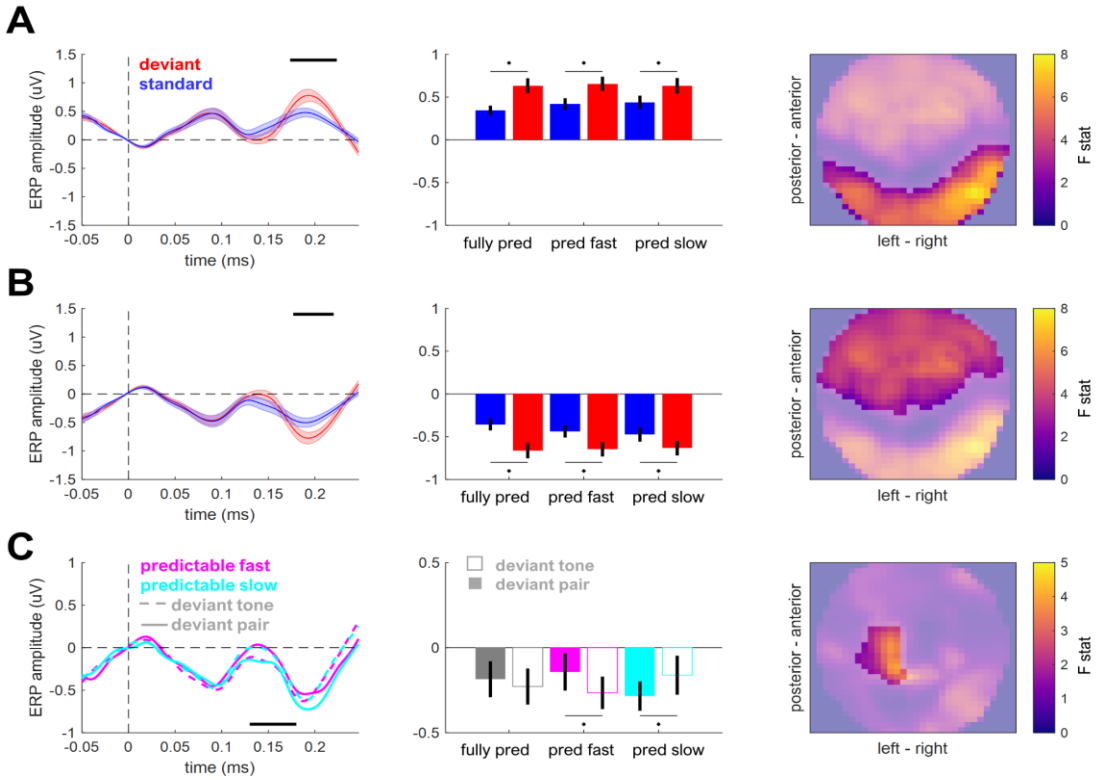

**Figure 18. Event-related potentials. (AB)** Main effect of content-based predictions (deviant vs. standard). Left panels: time courses of ERPs averaged over the spatial topography clusters shown in the right panels. Shaded area denotes SEM across participants. Black horizontal bar denotes $p_{FWE} < 0.05$. Middle panels: mean voltage values for standards (blue) and deviants (red). Right panels: spatial distribution of the main effect. Color bar: F value. **(C)** Hierarchical interaction between content-based predictions (deviant tone vs. deviant pair) and temporal predictions (predictable slow vs. predictable fast). Left panels: time courses of ERPs averaged over the spatial topography clusters shown in the right panels. Shaded area denotes SEM across participants. Black horizontal bar denotes $p_{FWE} < 0.05$. Middle panels: mean voltage values for the six deviant conditions. Right panels: spatial distribution of the interaction effect. Color bar: F value.

In the analysis of the main effect of "when" predictions (pooled over deviants and standards), no significant differences between the three temporal conditions were revealed (all $p_{FWE} > 0.05$). Similarly, in the analysis of the interaction effect of "what" and "when" predictions (pooled over deviant types), no significant effects were revealed. Specifically, neither deviants nor standards showed significant ERP amplitude differences when presented in different temporal contexts (all $p_{FWE} > 0.05$). Thus, the overall temporal structure of the sound sequences did not affect the tone-evoked responses (averaged across deviants and standards) or the mismatch responses (differences between deviants and standards).

However, an analysis of the interaction between "what" and "when" predictions focused on the hierarchical effects (i.e., differences between deviants presented in a corresponding temporal context, e.g. "fast" deviants in "predictable fast" condition, and those presented in a different context, e.g. "fast" deviants in "predictable slow" condition) revealed a significant interaction between deviant type and temporal condition (Fig. 18B; left central-posterior cluster: 130-180 ms, $F_{max} = 20.63$, $Z_{max} = 4.24$, $p_{FWE} = 0.044$). Post-hoc analysis revealed that MMR amplitudes in "predictable fast" were significantly larger for deviant tones (mean ± SEM: -0.1640 ± 0.0942 μV) than for deviant tone pairs (mean ± SEM: 0.0091 ± 0.1010 μV; $t_{19} = 2.2843$, $p = 0.0340$, two-tailed). In the "predictable slow" condition, MMR amplitude was observed to be nominally larger for deviant tone pairs (mean ± SEM: -0.1725 ± 0.0851 μV) than for deviant tones (mean ± SEM: -0.0155 ± 0.1233 μV), with this effect trending towards significance ($t_{19} = 1.9024$, $p = 0.0724$, two-tailed). No significant interaction effects were revealed when comparing deviant types between the "fully predictable" condition and either the "predictable slow" or the "predictable fast" conditions. Thus, we observed a specific increase in deviant ERP amplitude when this deviant was presented in a temporally congruent context.

***Brain-behavior correlation analysis***

Three neural predictors (the congruency effect of "what" and "when" predictions on

ERPs; the effect of "when" predictions on ITPC; and the effect of "what" predictions on ERPs) were tested as potential correlates of the behavioral effects on accuracy. We identified two outlier participants based on a linear regression model. Having excluded these two participants, we did not find any significant correlations between the neural indices and the behavioral effects (Pearson's *r;* all $p > 0.2$), suggesting that behavior in the repetition detection task is not functionally related to ERP signatures of deviance detection. However, we did find a significant correlation between the congruency effect on ERPs and the effect of "when" predictions on ITPC ($r = 0.6439$; $p = 0.0039$; corrected), such that the magnitude of the ERP difference between deviants presented in the temporally congruent vs. incongruent conditions positively correlated with the magnitude of the ITPC difference between "predictable fast" and "predictable slow" conditions.

***Source reconstruction***

To identify the most plausible sources underlying the observed ERP differences between deviants and standards, as well as the hierarchical interaction between deviant types and temporal conditions, we carried out a source reconstruction analysis. Overall, source reconstruction explained 76.43 ± 3.08% (mean ± SEM across participants) of sensor-level variance.

**Figure 19**

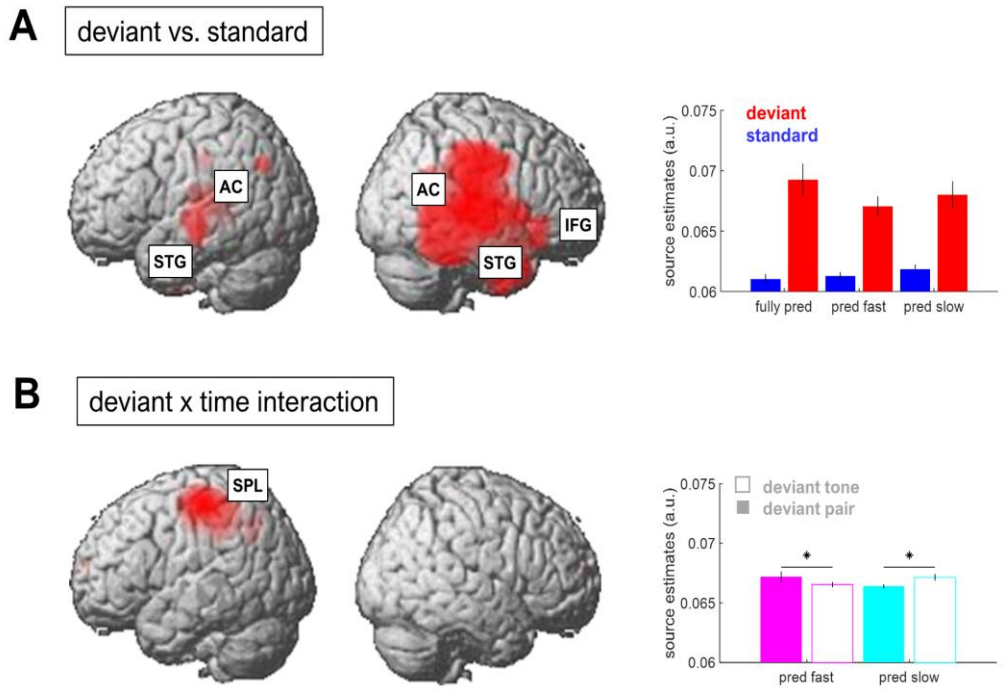


**Figure 19. Source reconstruction. (A)** Regions showing a significant main effect of content-based predictions (deviant vs. standard). Inset shows average source estimates per condition. Error bars denote SEM across participants. **(B)** Regions showing a significant hierarchical interaction effect between content-based predictions (deviant tone vs. deviant pair) and temporal predictions (predictable fast vs. predictable slow). Figure legend as in (A).

**Table 2**

| Effect | Cluster label | Peak MNI coords | $F_{max}$ | $Z_{max}$ | Voxel extent | $p_{FWE}$ |
|---|---|---|---|---|---|---|
| Deviant vs. standard | Right transverse temporal gyrus / auditory cortex (AC) | 48 -20 12 | 53.99 | 4.83 | 20508 | < 0.001 |
| | Right superior temporal gyrus (STG) | 44 -48 12 | 40.15 | 4.42 | | |
| | Right inferior frontal gyrus (IFG) | 40 26 -6 | 34.52 | 4.20 | | |
| | Left transverse temporal gyrus / auditory cortex (AC) | -38 -28 12 | 34.31 | 4.20 | 2177 | 0.003 |
| | Left superior temporal gyrus (STG) | -60 -20 -8 | 31.19 | 4.06 | | |
| ("Fast" vs. "slow" deviant) x ("Predictable fast" vs. "Predictable slow") | Left superior parietal lobule (lSPL) | -26 -40 46 | 49.37 | 5.82 | 3073 | 0.003 |

**Table 2. Source reconstruction results.** Summary of significant clusters showing differences between conditions.

The difference between source estimates associated with deviants and standards was localized to a large network of regions (see Table 2 for full results), including bilateral AC and STG and the rIFG. On the other hand, the interaction effect between deviant

types and temporal conditions was localized to a spatially confined cluster in the left superior parietal lobule (SPL; see Table 2). A post-hoc analysis revealed that, in this cluster, “fast” deviant responses presented in the “predictable fast” condition were associated with weaker source estimates than “slow” deviant responses presented in the “predictable fast” condition ($T_{max}$ = 3.67, $Z_{max}$ = 3.46, $p_{FWE}$ = 0.009, small-volume corrected). Similarly, “slow” deviant responses presented in the “predictable slow” condition were associated with weaker source estimates than “fast” deviant responses presented in the “predictable slow” condition ($T_{max}$ = 5.79, $Z_{max}$ = 5.11, $p_{FWE}$ = 0.003, small-volume corrected). Thus, while the deviant processing could be linked to a wide network of auditory and frontal regions, deviants presented in the corresponding temporal predictability conditions (e.g., “fast” deviants in the “predictable fast” context) were associated with a relative decrease of left parietal activity.

## Discussion

In the present study, we observed a correlation in MMR amplitude to heirachally-high (tone-pair) and hierarchically-low (single-tone) deviants in corresponding hierarchically structured slow and fast temporal predictability, indicating a congruency effect in predictive processing between so-called “what” and “when” events in the auditory system. This finding is supported by the presence of decreased left parietal

activity during congruent deviants. ITPC analysis revealed increased right hemispheric activity during the presence of slow temporal manipulations, consistent with existing research (Giroud et al., 2020).

Deviant responses within melodic sequences and tone contours are well documented, having been used to explore a variety of phenomena in the auditory system (see Yu et al., 2015 for partial review). Deviant tones within familiar musical scales have been found to elicit higher MMR amplitudes compared to those of unfamiliar scales and tones presented without a scale structure (Brattico et al., 2001), as well as higher deviant responses to out-of-scale notes in unfamiliar melodies (Brattico et al., 2006). Deviant responses to manipulated musical characteristics within melodic sequences (e.g. timing, pitch, transposition, melodic contour) have similarly been demonstrated in musician and non-musician groups (Tervaniemi et al., 2014, 2014; Vuust et al., 2011). In predictive coding frameworks, such evoked responses can be understood in the context of prediction error, wherein bottom-up error signaling triggers the adjustment of higher-level models of the stimulus train formed as a result of perceptual learning during repeated stimulus presentation (Garrido et al., 2009). Our source reconstruction was equally consistent with existing literature revealing bilateral activity in the primary auditory cortex (A1) and higher-order auditory regions in the superior temporal gyrus

(STG), as well as the rIFG (Garrido et al., 2008; Giroud et al., 2020).

Results of our frequency domain analysis show that the EEG spectrum largely follows that of the stimulus spectrum. However, ITPC peaks at the pair-tone rate of 'fully predictable' and 'predictable fast' are significantly larger than neighboring frequencies, which is not the case in the stimulus spectrum, indicating that ITPC peaks do not just follow the stimulus spectrum but also reflect the neural processing of sequence structures at higher levels (e.g. chunking). This result is inconsistent with previous MEG studies observing that neural entrainment at syllabic rates are bilaterally distributed (Ding et al., 2016). We found that the EEG-based ITPC response at tone-rate is stronger near central electrodes, with results consistent with existing EEG studies (Ding et al., 2017). Additionally, the pair-rate effect is predominantly present in the right hemisphere, suggesting that the hierarchical structure of non-linguistic sequences can be entrained by parallel neural activity in different regions at distinct time scales. Interestingly, the ITPC differences between conditions (predictable-low, predictable-fast) emerged during the experiment in pair-rate peaks, but not in tone-rate peaks, suggesting that rapid learning could enhance the congruence of neural entrainment to auditory sequences with different regularities at the pair-rate level. Similarly, a previous study (Moser et al., 2021) found significant differences in non-linguistic triplet-rate

ITPC peaks between structured and random conditions, occurring during early exposure. This ITPC difference suggests a fine shift in sequence encoding, with different regularities from single elements to integrated chunks. Notably, we also found correlations between the ITPC difference conditions and the congruence effect of ERP amplitude, indicating a mutual network between neural entrainment and prediction.

Previous studies have shown that the processing of musical information requires predictive mechanisms for timing of content of auditory events, and that such predictions can have modulatory effects at hierarchical cortical levels when presented within the framework of melodic expectation (Di Liberto et al., 2020; Royal et al., 2016). Musical stimuli present us with an intriguing opportunity to investigate predictive coding mechanisms, as the statistical regularities within musical frameworks are well defined and intrinsically learned. In particular, such structures allow us to disassociate “what” and “when” predictions while keeping other elements of a stimulus stream intact across manipulations and trials. Studies have demonstrated an early right anterior negativity (ERAN) in contexts where musical syntax has been violated, as opposed to the comparatively low-level acoustic diavations that elicit a MMN response (see Koelsch et al., 2019 for review). Because the presence of musical syntax violations require knowledge acquired through long-term repeated exposure to music, long-term

memory recall is involved in establishing those regularities. The role of memory in syntactical prediction violation is an avenue ripe for further investigation, and future studies may wish to extend our paradigm to further probe the observed late-series ITPC pair-rate differences in that context.

## References

Auksztulewicz, R., Friston, K., 2016. Repetition suppression and its contextual determinants in predictive coding. Cortex, Special Issue:Repetition suppression-an integrative view 80, 125–140. https://doi.org/10.1016/j.cortex.2015.11.024

Auksztulewicz, R., Friston, K., 2015. Attentional Enhancement of Auditory Mismatch Responses: a DCM/MEG Study. Cereb. Cortex 25, 4273–4283. https://doi.org/10.1093/cercor/bhu323

Auksztulewicz, R., Myers, N.E., Schnupp, J.W., Nobre, A.C., 2019. Rhythmic Temporal Expectation Boosts Neural Activity by Increasing Neural Gain. J. Neurosci. 39, 9806–9817. https://doi.org/10.1523/JNEUROSCI.0925-19.2019

Auksztulewicz, R., Schwiedrzik, C.M., Thesen, T., Doyle, W., Devinsky, O., Nobre, A.C., Schroeder, C.E., Friston, K.J., Melloni, L., 2018. Not All Predictions Are Equal: "What" and "When" Predictions Modulate Activity in Auditory Cortex through Different Mechanisms. J. Neurosci. 38, 8680–8693. https://doi.org/10.1523/JNEUROSCI.0369-18.2018

Belardinelli, P., Ortiz, E., Braun, C., 2012. Source Activity Correlation Effects on LCMV Beamformers in a Realistic Measurement Environment. Comput. Math. Methods Med. 2012, e190513. https://doi.org/10.1155/2012/190513

Brattico, E., Näääätäänen, R., Tervaniemi, M., 2001. Context Effects on Pitch

Perception in Musicians and Nonmusicians: Evidence from Event-Related-Potential Recordings. Music Percept. 19, 199–222. https://doi.org/10.1525/mp.2001.19.2.199

Brattico, E., Tervaniemi, M., Näätänen, R., Peretz, I., 2006. Musical scale properties are automatically processed in the human auditory cortex. Brain Res. 1117, 162–174. https://doi.org/10.1016/j.brainres.2006.08.023

de Cheveigné, A., Simon, J.Z., 2008. Denoising based on spatial filtering. J. Neurosci. Methods 171, 331–339. https://doi.org/10.1016/j.jneumeth.2008.03.015

Denham, S.L., Winkler, I., 2020. Predictive coding in auditory perception: challenges and unresolved questions. Eur. J. Neurosci. 51, 1151–1160. https://doi.org/10.1111/ejn.13802

Di Liberto, G.M., Pelofi, C., Bianco, R., Patel, P., Mehta, A.D., Herrero, J.L., de Cheveigné, A., Shamma, S., Mesgarani, N., 2020. Cortical encoding of melodic expectations in human temporal cortex. eLife 9, e51784. https://doi.org/10.7554/eLife.51784

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., Poeppel, D., 2017. Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG). Front. Hum. Neurosci. 11, 481. https://doi.org/10.3389/fnhum.2017.00481

Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of

hierarchical linguistic structures in connected speech. Nat. Neurosci. 19, 158–164. https://doi.org/10.1038/nn.4186

Ding, N., Simon, J.Z., 2013. Power and phase properties of oscillatory neural responses in the presence of background activity. J. Comput. Neurosci. 34, 337–343. https://doi.org/10.1007/s10827-012-0424-6

Emmendorfer, A.K., Correia, J.M., Jansma, B.M., Kotz, S.A., Bonte, M., 2020. ERP mismatch response to phonological and temporal regularities in speech. Sci. Rep. 10, 9917. https://doi.org/10.1038/s41598-020-66824-x

Fitzgerald, K., Auksztulewicz, R., Provost, A., Paton, B., Howard, Z., Todd, J., 2021. Hierarchical Learning of Statistical Regularities over Multiple Timescales of Sound Sequence Processing: A Dynamic Causal Modeling Study. J. Cogn. Neurosci. 33, 1549–1562. https://doi.org/10.1162/jocn_a_01735

Friston, K., Penny, W., 2011. Post hoc Bayesian model selection. NeuroImage 56, 2089–2099. https://doi.org/10.1016/j.neuroimage.2011.03.062

Friston, K., Zeidman, P., Litvak, V., 2015. Empirical Bayes for DCM: A Group Inversion Scheme. Front. Syst. Neurosci. 9, 164. https://doi.org/10.3389/fnsys.2015.00164

Garrido, M.I., Friston, K.J., Kiebel, S.J., Stephan, K.E., Baldeweg, T., Kilner, J.M., 2008. The functional anatomy of the MMN: A DCM study of the roving paradigm. NeuroImage 42, 936–944.

https://doi.org/10.1016/j.neuroimage.2008.05.018

Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009. The mismatch negativity: A review of underlying mechanisms. Clin. Neurophysiol. 120, 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Giroud, J., Trébuchon, A., Schön, D., Marquis, P., Liegeois-Chauvel, C., Poeppel, D., Morillon, B., 2020. Asymmetric sampling in human auditory cortex reveals spectral processing hierarchy. PLOS Biol. 18, e3000207. https://doi.org/10.1371/journal.pbio.3000207

Haegens, S., Zion Golumbic, E., 2018. Rhythmic facilitation of sensory processing: A critical review. Neurosci. Biobehav. Rev. 86, 150–165. https://doi.org/10.1016/j.neubiorev.2017.12.002

Hasson, U., Chen, J., Honey, C.J., 2015. Hierarchical process memory: memory as an integral component of information processing. Trends Cogn. Sci. 19, 304–313. https://doi.org/10.1016/j.tics.2015.04.006

Ille, N., Berg, P., Scherg, M., 2002. Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc. 19, 113–124. https://doi.org/10.1097/00004691-200203000-00002

Jalewa, J., Todd, J., Michie, P.T., Hodgson, D.M., Harms, L., 2020. Do rat auditory event related potentials exhibit human mismatch negativity attributes related to

predictive coding? Hear. Res. 107992. https://doi.org/10.1016/j.heares.2020.107992

Kawahara, H., 2006. STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. Acoust Sci Technol 27349. https://doi.org/10.1250/ast.27.349

Kilner, J.M., Kiebel, S.J., Friston, K.J., 2005. Applications of random field theory to electrophysiology. Neurosci. Lett. 374, 174–178. https://doi.org/10.1016/j.neulet.2004.10.052

Koelsch, S., Vuust, P., Friston, K., 2019. Predictive Processes and the Peculiar Case of Music. Trends Cogn. Sci. 23, 63–77. https://doi.org/10.1016/j.tics.2018.10.006

Little, S., Bonaiuto, J., Meyer, S.S., Lopez, J., Bestmann, S., Barnes, G., 2018. Quantifying the performance of MEG source reconstruction using resting state data. NeuroImage 181, 453–460. https://doi.org/10.1016/j.neuroimage.2018.07.030

Litvak, V., Friston, K., 2008. Electromagnetic source reconstruction for group studies. NeuroImage 42, 1490–1498. https://doi.org/10.1016/j.neuroimage.2008.06.022

Moser, J., Batterink, L., Li Hegner, Y., Schleger, F., Braun, C., Paller, K.A., Preissl, H., 2021. Dynamics of nonlinguistic statistical learning: From neural entrainment to the emergence of explicit knowledge. Neuroimage 240, 118378. https://doi.org/10.1016/j.neuroimage.2021.118378

Musacchia, G., Large, E., Schroeder, C.E., 2014. Thalamocortical mechanisms for integrating musical tone and rhythm. Hear. Res. 308, 50–59. https://doi.org/10.1016/j.heares.2013.09.017

Rosch, R.E., Auksztulewicz, R., Leung, P.D., Friston, K.J., Baldeweg, T., 2019. Selective Prefrontal Disinhibition in a Roving Auditory Oddball Paradigm Under N-Methyl-D-Aspartate Receptor Blockade. Biol. Psychiatry Cogn. Neurosci. Neuroimaging 4, 140–150. https://doi.org/10.1016/j.bpsc.2018.07.003

Royal, I., Vuvan, D.T., Zendel, B.R., Robitaille, N., Schönwiesner, M., Peretz, I., 2016. Activation in the Right Inferior Parietal Lobule Reflects the Representation of Musical Structure beyond Simple Pitch Discrimination. PLoS ONE 11, e0155291. https://doi.org/10.1371/journal.pone.0155291

Schroeder, C.E., Lakatos, P., 2009. Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. 32, 9–18. https://doi.org/10.1016/j.tins.2008.09.012

Tervaniemi, M., Huotilainen, M., Brattico, E., 2014. Melodic multi-feature paradigm reveals auditory profiles in music-sound encoding. Front. Hum. Neurosci. 8, 496. https://doi.org/10.3389/fnhum.2014.00496

Todorovic, A., Auksztulewicz, R., 2021. Dissociable neural effects of temporal expectations due to passage of time and contextual probability. Hear. Res. 399, 107871. https://doi.org/10.1016/j.heares.2019.107871

Vuust, P., Brattico, E., Glerean, E., Seppänen, M., Pakarinen, S., Tervaniemi, M., Näätänen, R., 2011. New fast mismatch negativity paradigm for determining the neural prerequisites for musical ability. Cortex 47, 1091–1098. https://doi.org/10.1016/j.cortex.2011.04.026

Wipf, D., Nagarajan, S., 2009. A unified Bayesian framework for MEG/EEG source imaging. NeuroImage 44, 947–966. https://doi.org/10.1016/j.neuroimage.2008.02.059

Yu, X., Liu, T., Gao, D., 2015. The Mismatch Negativity: An Indicator of Perception of Regularities in Music. Behav. Neurol. 2015, 469508. https://doi.org/10.1155/2015/469508

# Chapter 5. Conclusions

This thesis investigated the predictive processing and sequence segmentation ability in humans and rodents. We provided more evidence for the potential mechanisms of acoustic sequence coding in different species and for the processing hierarchy in the auditory neocortex. The main findings can be summarized as follows:

## 1. Learning boosts the decoding of sound sequences in rat auditory cortex

(Ding et al., 2016) found that human subjects' cortical activity is entrained to distinct time scales of speech streams across different linguistic hierarchical levels if subjects are familiar with the language. In our study, we sought to establish an animal model of entrainment to different hierarchical levels of acoustic sequences, and tested whether neural activity in the rat auditory cortex is modulated by previous segment experience (familiarity).

1.1 We found that rats could behaviorally differentiate familiar from unfamiliar acoustic stimulus triplets after training. However, the low-frequency entrainment to the triplet rate was not detectable in the neural activity of anesthetized rodents.

1.2 We showed that behavioral training leads to improvements in decoding stimulus-related information for familiar sequences based on the spatial pattern of neural activity in the auditory cortex.

Thus, our results suggest that activity in the rat auditory cortex is modulated by previous stimulus experience and enables improved decoding of familiar sound sequences.

## 2. Deviant Processing of Complex Sounds in Mouse Auditory Cortex

Extensive literature in rodent models (Parras et al., 2021; Parras et al., 2017) on mismatch responses to single stimuli suggests a hierarchy of regions, such that secondary areas encode deviance detection to a larger extent than primary areas. However, it remains elusive whether the primary auditory cortex, or higher order regions, can also encode deviance of more complex stimulus patterns. This study investigated whether mismatch responses to complex deviants differ between primary and higher-order auditory regions.

2.1 Our results suggest that mice could encode both the local probabilities (transitions) and the more global stimulus patterns, and elicit mismatch signals to substitution deviants (pairs containing novel elements) and transposition

deviants (pairs containing familiar elements in a different order), but not to omission deviants.

2.2 Notably, the higher-order area of the auditory cortex (A2) elicited more pronounced mismatch responses to complex deviants than the primary auditory area.

Thus, our results suggest a hierarchical gradient of prediction error signaling for complex deviants.

## 3. Congruence of "What" and "When" Predictions in the Auditory System

Transitions and timing knowledge is one of the most basic mechanisms of sequence coding (Dehaene et al., 2015). A recent human ECoG study (Auksztulewicz et al., 2018) found that cortical responses to auditory targets were modulated both by the stimulus content expectation and by its timing expectation, and suggested that what (content) and when (time) predictions engage complementary and interacting neural mechanisms in different cortical regions. Here, we tested whether “what” and “when” predictions show interactions specific to the hierarchical level at which they are manipulated (e.g., single tones vs. tone pairs) by independently manipulating temporal and

content-based expectations at two hierarchical levels in humans.

3.1 The EEG experiment in humans revealed interactive effects of “what ” and “when” predictions in terms of their modulation of the stimulus-evoked activity amplitude. This effect was only present at both hierarchical levels (single tones and tone pairs) only for hierarchically congruent “what” and “when” predictions (i.e., fast predictions modulated mismatch responses to unexpected single tones, and slow predictions modulated mismatch responses to unexpected tone pairs)

3.2 Frequency-domain analysis showed that entrainment to stimuli expected in time at the slower temporal scale gradually increased over the course of the experiment and was mostly expressed over the right hemisphere.

Thus, our results indicated a congruence effect in predictive processing of stimulus contents and timing in the auditory system.

## 4. Limitations and future perspectives

While the work in this thesis has advanced our understanding, it nevertheless leaves some questions unanswered, and there are aspects of this work which, with the benefit of hindsight, could be improved. For example, in the first study (chapter 2) the training method could be improved in future studies to shorten the training duration,

which can also avoid the potential influence of the age effect. However, the rats were normal hearing, and there were no cognitive tasks (Rimmele et al., 2015) during the ECoG recording, which reduced the potential confound of age effects in our first study. Our results (chapter 2) suggest that rats process suprasyllabic regularities in acoustic sequences, and that the training experience modulates their cortical activity even under subsequent anesthesia. However, we didn't observe the neural entrainment at the triplet rate in the trained animal group and we cannot exclude that the animal status may strongly affect the neural oscillations (Akeju et al., 2017) in our study. Previous studies in humans suggested that neural entrainment could be disrupted by sleep (Makov et al., 2017). In addition, different anesthesia methods (Carbajal et al., 2018, Schmidt et al., 2013) may influence the neural response in animals differently, especially for the MMN. Thus, future studies should test the behavioral relevance of these signals by relating the neural activity to behavioral responses in awake and behaving animals.

The second study's experimental design could be improved. For example, we have transposition (more global) and substitution (position) of deviant stimuli, and we also introduced the novelty (probability) of the deviant stimuli. This design makes it hard to explain which factor (novelty or position) modulates the neural response most. Future studies should split these factors into different experiments or blocks. For the stimulus selection, especially for the omission condition, future studies should employ simpler

stimuli, such as single vowels instead of syllable pairs, to detect the omission response.

In addition, while we demonstrate that both mice and humans could encode the transitions and temporal information of the sequence (chapter 3 and 4), future research should employ the same paradigm in both species to better identify the possible scope of extrapolating results from animal models to humans.

Nevertheless, the research presented in this thesis indicates that two building blocks of sequence coding mechanisms (i.e., transitions and timing, and chunking; Dehaene et al., 2015) are evolutionarily conserved across rodents and humans.

## References


Auksztulewicz, R., Schwiedrzik, C. M., Thesen, T., Doyle, W., Devinsky, O., Nobre, A. C., Schroeder, C. E., Friston, K. J., & Melloni, L. (2018). Not all predictions are equal:“What” and “when” predictions modulate activity in auditory cortex through different mechanisms. *Journal of Neuroscience, 38*(40), 8680-8693.

Akeju, O., & Brown, E. N. (2017). Neural oscillations demonstrate that general anesthesia and sedative states are neurophysiologically distinct from sleep. Current opinion in neurobiology, 44, 178-185.

Carbajal, G. V., & Malmierca, M. S. (2018). The neuronal basis of predictive coding along the auditory pathway: from the subcortical roots to cortical deviance detection. Trends in Hearing, 22, 2331216518784822.Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015, Oct 7). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron, 88*(1), 2-19. https://doi.org/10.1016/j.neuron.2015.09.019

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature neuroscience, 19*(1), 158-164.

Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., Golumbic, E.Z., 2017. Sleep disrupts high-level speech parsing despite significant basic auditory processing. J. Neurosci. 37(32), 7772-7781.

Parras, G. G., Casado-Román, L., Schröger, E., & Malmierca, M. S. (2021). The posterior auditory field is the chief generator of prediction error signals in the auditory cortex. *Neuroimage, 242*, 118446.

Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdes-Baizabal, C., Escera, C., & Malmierca, M. S. (2017, Dec 15). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nat Commun, 8*(1), 2148. https://doi.org/10.1038/s41467-017-02038-6

Rimmele, J. M., Sussman, E., & Poeppel, D. (2015). The role of temporal structure in the investigation of sensory memory, auditory scene analysis, and speech

perception: A healthy-aging perspective. *International Journal of Psychophysiology*, 95(2), 175-183.

Schmidt, A., Diaconescu, A. O., Kometer, M., Friston, K. J., Stephan, K. E., & Vollenweider, F. X. (2013). Modeling ketamine effects on synaptic plasticity during the mismatch negativity. Cerebral cortex, 23(10), 2394-2406.