Uncovering the Brain Mechanisms that Underlie the Detection of Patterns in Sound Sequences

Mingyue Hu

This thesis was submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Ear institute, Faculty of Brain Sciences, University College London, UK

July 2024

Supervisor:

Professor Maria Chait

Declaration

I, Mingyue Hu, confirm that the work presented in this thesis is solely my own. Where information has been derived from other sources, appropriate citations have been provided in the thesis. Additionally, artificial intelligence has been utilised to correct grammatical errors and refine some sentence structures.

Mingyue Hu

July 2024

Abstract

The auditory system is wired to be sensitive to patterns in sound, this ability is crucial for the brain to comprehend and interact with its environment. Despite the pivotal role of these processes, the underlying neural mechanisms remain poorly understood. This lack of clarity hinders the development of comprehensive models of auditory processing, machine intelligence, and the creation of targeted interventions for those clinical diseases involving auditory impairment.

Expanding on those unsolved issues, this PhD thesis explores several questions surrounding the cognitive processes and neural mechanisms underpinning the auditory system's sensitivity to patterns in human listeners. Chapter 2 investigates the constraints of auditory memory in pattern recognition. In this behavioural study, participants engaged in identifying emerging patterns within rapidly unfolding sound sequences. By varying the durations and informational complexities of patterns, the study assessed whether memory integration is primarily duration-dependent or also configured to monitor discrete items. Chapter 3 explored the neural mechanisms that support pattern detection. Participants passively listened to predictable and unpredictable sound sequences while their brain responses were recorded with magnetoencephalography (MEG). By analysing the time domain signal and localising the neural sources, the study is investigating how the brain is representing the predictable sensory signals. Chapter 4 addresses the challenges in detecting patterns when sound sequence is slowed down, examining how auditory short-term memory, sustained attention, frequency discrimination, and task engagement influence detection performance. Building on the findings of Chapter 4, Chapter 5 employed Electroencephalography (EEG) to further explore the correlation between short-term memory abilities and implicit pattern detection. Participants listened passively to sound patterns of 5500ms, and both EEG signals and behavioural performance data were analysed.

Overall, the results from this thesis align with predictive coding theory and provide novel insights into the neural underpinnings of perceptual and cognitive processes underlying auditory pattern detection.

Impact Statement

This PhD thesis enhances the field of cognitive neuroscience by exploring the neural mechanisms that underpin the auditory system's sensitivity to patterns in healthy and young human listeners. It broadens our understanding in three key areas: the constraints of auditory memory in recognizing patterns, the neural foundations that support pattern detection, and the influence of cognitive factors on pattern detection efficacy. By elucidating the mechanisms behind these processes, this project provides evidence that could be crucial in addressing challenges associated with disorders such as schizophrenia, autism, and other mental or neurological diseases that manifest with auditory processing disorders and memory deficits.

The study challenges existing auditory models by demonstrating through behavioral experiments that pattern recognition is not solely dependent on the duration of sounds but is also intricately associated with the item-wise informational content within auditory sequences. This insight shifts the theoretical model from simply monitoring temporal information at a fixed pace to a more nuanced understanding that the brain adaptively integrates discrete elements within sound sequences. These findings have significant implications for developing more effective auditory and memory models and provide valuable insights into understanding auditory and memory disorders.

In investigating the neural mechanisms of pattern detection, the thesis examines how the brain processes sound sequences that are either predictable or unpredictable by introducing silent gap between tones. The MEG findings align with predictive coding theory, revealing the coexistence of dual neural components involved in analysing auditory inputs. The sustained response, supported by a neural network encompassing the auditory cortex, hippocampus, and inferior frontal cortex, is associated with the precision. In contrast, the phasic response evoked by individual tones, primarily involving the auditory cortex and inferior frontal cortex, carries information of prediction errors. These results provide novel insights into how these multiplexed neural processes collectively shape the representations of auditory pattern.

Lastly, the exploration of cognitive factors impacting pattern detection performance reveals a correlation between the sustained neural response elicited by sound patterns and short-term memory abilities. This discovery, the first from this thesis, provided novel insights suggesting shared neural pathways between implicit pattern detection and auditory short-term memory. The hippocampus is hypothesized to be the crucial area where this overlap occurs, setting the stage for future investigations on the nature of sustained response and the neural mechanism of auditory scene analysis.

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

1. For the Chapter 3 (Study 2) that has already been published:

(a) What is the title of the manuscript?

The title of the manuscript is 'Concurrent Encoding of Sequence Predictability and Event-Evoked Prediction Error in Unfolding Auditory Patterns'.

(b) Please include a link to or doi for the work:

The doi of the manuscript: https://doi.org/10.1523/JNEUROSCI.1894-23.2024

(c) Where was the work published?

The Journal of Neuroscience

(d) Who published the work?

The senior editor: Anna Nobre

The reviewing editor: Jonas Obleser

The editor-in-chief: Sabine Kastner

(e) When was the work published?

It was published on April 3, 2024

(f) List the manuscript's authors in the order they appear on the publication:

Mingyue Hu, Roberta Bianco, Antonio Rodriguez Hidalgo, Maria Chait

(g) Was the work peer reviewed?

Yes, it was peer reviewed.

(h) Have you retained the copyright?

Yes, I have retained the copyright.

(i) Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)? If 'Yes', please give a link or doi; If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

Yes, the doi is: https://doi.org/10.1101/2023.10.06.561171

⊗ I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.

Acknowledgement

Upon completing my thesis, I extend my deepest gratitude to those who supported and guided me through my four-year PhD journey.

First and foremost, I am profoundly grateful to my supervisor, Prof. Maria Chait, for guiding me into this captivating field. Her innovative ideas, broad perspective and passion for research have greatly shaped my understanding of neuroscience and how to be a scientist. Due to the pandemic, my PhD project was adjusted accordingly, and the challenges we faced were substantial. Most projects must be developed quickly, and we had to deal with a lot of uncertainties. Nevertheless, my supervisor's extensive support made me feel that I was not facing these challenges alone. Her adept management of the lab and projects ensured stability over the course of my PhD. I am deeply thankful for her leadership during that crucial period.

I am also profoundly grateful to Dr. Roberta Bianco, who introduced me to technical measurement and analysis methods from the very beginning. Throughout these years, her expert guidance, willingness to share her experience, and encouragement have been invaluable.

Moreover, my sincere thanks go to Dr. Alex Billig for his advice in data analysis and data interpretation, and to Dr. Alice Milne for her expert assistance on programming.

I also want to express my heartfelt appreciation to Dr. Mercede Erfanianghasab, Claudia Contadini-Wright, Kaho Magami, Mert Huviyetli, and Buse Adam for their companionship at UCL. Their presence shielded me from loneliness and gave me treasured, warm memories.

Lastly, I extend my gratitude to my parents for their unwavering care and unconditional love from thousands of miles away. Their support in all my endeavours has been invaluable. I also wish to thank my boyfriend, Bob, for his constant encouragement and steadfast support during the toughest periods of my PhD. Their presence has been a cornerstone of my success.

Table of Contents

Declaration 1

Abstract 2

Impact Statement 3

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s) 4

Acknowledgement 5

Table of Contents 6

Table of Figures 10

Chapter 1: General Introduction 12

1.1. Comprehending Environments by Using Sounds 12

1.1.1. How do we define sound? 12

1.1.2. The Role of Auditory Objects in Understanding the Environment 12

1.1.3. Auditory Cues in Complex Acoustic Environments 13

1.1.4. Modelling Auditory Scene Using Pure Tones 16

1.2. Predictive Coding Theory 17

1.2.1. Bayesian Processing in Perception 18

1.2.2. Hierarchical Models of Predictive Coding 20

1.2.3. Neural Evidence of Prediction Error 22

1.2.4. What is Precision 26

1.3. Sensitivity to Patterns 28

1.3.1. Statistical Learning in Auditory Scene 28

1.3.2. The Neural Underpinnings of Statistical Learning 30

1.3.3. Neural Correlates of Auditory Regularity Encoding 34

1.3.4. The Neural Substrates of Auditory Regularity Encoding 38

1.4. Aim of This Project 42

Chapter 2: Unravelling the Interplay of Duration and Information boundaries in Rapidly Unfolding Sound Pattern Detection: Insights from Behavioural Examination 43

2.1. Introduction 43

2.1.1. The Motivation Behind the Study 48

2.1.2. Aim & Hypotheses 53

2.2. Experiment 1. Is Pattern Detection Ability Limited by Pattern Complexity 55

2.2.1. Methods 55
2.2.2. Results 61
2.3. Experiment 2. How is Pattern Detection Performance Affected by Increasing Pattern Duration 64
2.3.1. Methods 65
2.3.2. Results 66
2.4. Experiment 3 69
2.4.1. Methods 70
2.4.2. Results 72
2.5. Discussion 73
2.5.1. Response Times Reflect Dynamic Information Processing in Auditory Memory 74
2.5.2. Increased Informational Complexity Facilitate Pattern Detection but not Behave Like an Ideal Observer Model 75
2.5.3. Dynamic Nature of Auditory Memory 76
2.5.4. Does the Brain Process Fast and Slow Sounds Differently 77
2.5.5 Conclusion and Future Direction 79

Chapter 3: Concurrent Encoding of Precision and Event-evoked Prediction Error in Unfolding Auditory Patterns 81
3.1. Introduction 81
3.2. Methods 83
3.2.1. Experiment 1 - Online Behavioural Study 83
3.2.2. Experiment 2 - MEG in Naïve Passively Listening Participants 86
3.3. Results 92
3.3.1. Behavioural Performance Reveals Good Sensitivity to Regularity Even Following the Introduction of Silent Gaps between Tones. 92
3.3.2. The Emergence of Regularity is Associated with an Increase in Sustained MEG Activity 92
3.3.3. No Significant Correlation Between Tone-evoked and Sustained-response Effects 101
3.4. Discussion 101
3.4.1. Sustained Brain Responses Track Pattern Emergence Even in Slow Sequences 101
3.4.2. Reduced Responses to Tones in REG Relative to RND Patterns 103
3.4.3. Multiplexed Representation of Sequence Predictability 104
3.5. Supplementary Information 106

Chapter 4. Whether/which Cognitive Factor Might Account for the Variability in Pattern Detection Performance. 107

Sub-study 1: Comparison Between Auditory SART and Visual SART 108

4.1. Introduction 108

4.2. Methods 110
4.2.1. Visual SART 110
4.2.2. Auditory SART 110
4.2.3. Participants 111

4.3. Results 111
4.3.1. NoGo Commission Errors 112
4.3.2. Go Trial Response Times 112
4.3.3. NoGo Trial Commission Error Rate and Go Trial Response Times 113

4.4. Discussion 116
4.4.1. What Contributed to the Prolonged Response Times Observed in Auditory SART Go Trials 116
4.4.2. What does the Commission Error Reflect 116
4.4.3. Why Reduced Commission Errors were Performed in Auditory SART 117
4.4.4. Visual SART or Auditory SART 118

Sub-study 2: Cognitive Underpinnings of Auditory Pattern Detection 119

4.5. Introduction 119
4.5.1. Auditory Working/short-term Memory 120
4.5.2. Sustained Attention 121
4.5.3. General Task Engagement and Vigilance 122

4.6. Methods 122
4.6.1. Participants 122
4.6.2. Pattern Detection Task 123
4.6.3. Visual SART 124
4.6.4. Frequency Sensitivity Test (FST) 124
4.6.5. Tone Pattern Comparison Test (TP-COMP) 124

4.7. Results 125

4.8. Discussion 132
4.8.1. Sustained Attention and Frequency Discrimination's Role in Gap100 Task Performance 132
4.8.2. Auditory Short-Term Memory Predicts Performance in Explicit Sound Pattern Detection Across Two Time Scales 134
4.8.3. Implications for Neural Mechanism 135

4.9. Conclusion 136

Chapter 5: Sensitivity to Complex Sound Patterns is Correlated with Auditory Short Term Memory 137

5.1. Introduction 137
5.1.1. The Goal of the Study 139
5.2. Methods: 141
5.2.1. Experiment 141
5.2.2. Data Analysis 142
5.2.3. Tone Pattern Comparison Task (TP-COMP): 144
5.3. Results: 147
5.3.1. Sequence-Evoked EEG Responses Suggest Regularity Extraction 147
5.3.2. Performance in Tone Pattern Comparison Task exhibit Significant Variability 148
5.3.3. TP-COMP Performance Associated with Differential EEG Responses to REG Patterns 148
5.3.4. Tone-Evoked Responses Indicate Regularity Encoding 152
5.3.5. Inter-Group Comparisons of Phasic Responses Highlight Variations in Auditory Processing of Individuals 156
5.3.6. Enhanced Early Auditory Response was Observed in Memory Task High Performers 159
5.3.7. Correlation Between Tone-evoked Activity and Sustained Response 161
5.4. Discussion: 163
5.4.1. Shared Neural Processes Between Implicit Sound Pattern Detection and Auditory Short-term Memory 163
5.4.2. Neural Processes Underlying N2 are Associated with Regularity Encoding 165
5.4.3. Correlation Between Sequence Evoked Sustained Response and Tone Evoked Phasic Activity 167
5.4.4. Temporal Adaptation in N1 168
5.5. Conclusion 169

General Discussion 171
6.1. Summary of Main Findings 171
6.2. Implications for the Brain Functions 172
6.3. Limitations and Future Directions 173
6.3.1. Limitations 173
6.3.2. Future Direction 174

Reference 176
Author Contribution 199

Table of Figures

Figure 1.1. Brain response: The graph depicts the root mean square (RMS) of brain responses for pattern condition REG5, REG10, and REG15, as well as for the random sequence condition RAND20. 36

Figure 2.1. Group root mean square (RMS) of brain responses to REG5, REG10, and REG15 conditions, along with RAND20, are displayed. 47

Figure 2.2. In the context of PPM decay model (see methods section), this study examined how the echoic memory buffer size can affect the information content dynamics over time. 50

Figure 2.3. This demo illustrates how a pattern becomes theoretically detectable throughout the unfolding of a sound sequence over time (this example plots the spectrogram of RANDREG5), assuming the limited temporal capacity of the memory buffer as proposed by the PPM decay model. 51

Figure 2.4. Comparison of the model output between REG10 and REG20 in different temporal capacities of the echoic memory buffer. 52

Figure 2.5. Experiment 1 hypotheses.. 55

Figure 2.6. Spectrogram of example RANDREG stimuli of all Rcyc conditions in the short pattern duration (500ms). 58

Figure 2.7. Behavioural performance.. 62

Figure 2.8. Response times fall between the boundaries of two hypotheses. 64

Figure 2.9. d Prime distribution from experiment 2. 67

Figure 2.10. The results from experiment 1 were reproduced in experiment 2. 68

Figure 2.11. Individual distribution of response time across all conditions in Experiment 2.. 68

Figure 2.12. Individual $RT_{in\ ms}$ were measured in conditions REG10 and REG20 with a pattern duration of 500ms (Cyc500) in experiments 1, 2, and 3. 71

Figure 2.13. Individual $RT_{in\ ms}$ measured in condition REG10 and REG20 with pattern duration of 1500ms (Cyc1500) from experiment 2 and experiment 3. 72

Figure 3.1. Behavioural experiment. 86

Figure 3.2. Examples of stimuli in the MEG experiment. 89

Figure 3.3. MEG response to 'fast' (Gap0) sequences. 93

Figure 3.4. MEG response to 'slow' (Gap200) sequences. 95

Figure 3.5. Tone evoked responses. 97

Figure 3.6. The source analysis on the sequence evoked response on 'slow' conditions 106

Figure 4.1. Distribution of commission errors (hits on Nogo trials) across all NOGO trials for each of auditory, visual version of Sustained Attention Response Task (SART). 111

Figure 4.2. Significant correlation was seen across modalities in subjects group whose NoGo commission error rate was below 80% in visual SART (N=19). 113

Figure 4.3. The spearman correlation analysis was applied on both tasks, all individuals were included (N=24). 113

Figure 4.4. Coefficient of variation of response time (RTCV). 114

Figure 4.5. Spearman correlation. 115

Figure 4.6. The spectrogram of the stimuli in the tone pattern comparison task (TP-COMP). 125

Figure 4.7. Performance(d') and $RT_{number\ of\ tones}$ to pattern detection tasks for all participants (N = 109). 127

Figure 4.8. The simulation of pattern detection task performance modelled on the size 109 subjects. 128

Figure 4.9. Substantial variability was observed from all cognitive measurements. 129

Figure 4.10. Large variability in Gap500 task were still observed despite the fact that all subjects can achieve ceiling peformance in Gap100 task. 131

Figure 4.11. Spearman Correlation. 132

Figure 5.1. Brain response. 146

Figure 5.2. Memory task performance. 150

Figure 5.3. Sequence evoked EEG responses grouped based on TP-COMP performance. 152

Figure 5.4. Tone response across each cycle of REG and corresponding timing in RND. 155

Figure 5.5. Tone response comparison between TPCT above median performers and below median performers. 158

Figure 5.6. Tone-evoked activity in cycle 1. 160

Figure 5.7. Scatter plots for neural responses in the 7.5-11 s (Cycle 2) and the 11-16.5 s (Cycle 3) time windows. 162

1. Chapter 1: General Introduction

1.1. Comprehending Environments by Using Sounds

1.1.1. How do we define sound?

"If a tree falls in a forest and no one is around to hear it, does it make a sound?"

The Chautauquan in 1883

1.1.2. The Role of Auditory Objects in Understanding the Environment

In our everyday environment, we are constantly surrounded by overlapping sound signals. From birdsong to the hustle and bustle of city streets, our auditory system must navigate this complex auditory landscape and extract meaningful sounds to make sense of our surroundings and guide our behaviour. This complex process is called auditory scene analysis.

Auditory objects are central to auditory scene analysis which are typically emitted by specific sources as a consequence of physical actions. Although its definition is still under debate (Kubovy and Van Valkenburg, 2001; Griffiths and Warren, 2004; Dyson and Ishfaq, 2008; Shamma, 2008; Winkler et al., 2009; Moore et al., 2010; Schnupp et al., 2013), most studies summarize it as perceptual outcomes derived from the auditory system's ability to detect, segregate, and group spectrotemporal patterns in sound. More intuitively, these objects represent the auditory system's mental description of a potential source performing certain actions, thus generating stable perceptual units within the complex acoustic environment. (Bregman, 1990; Kubovy and Van Valkenburg, 2001; Shinn-Cunningham, 2008; Winkler et al., 2009; Winkler and Denham, 2024). This ability is particularly vital in environments with multiple sound sources, such as a forest with various animal calls, a busy street with cars ringing and people walking, or a crowded room with multiple conversations.

Each distinct sound or cluster of sounds is perceived as separate within an auditory scene constitutes an auditory object. Understanding how auditory objects are formed helps explain how people interact effectively with their environment.

In all sensory modalities, the concept of objects serves a similar ecological purpose: to quickly identify environmental sources and understand their actions. However, the ways in which these sensory inputs are processed vary significantly depending on the physical nature of the stimuli involved. For example, in vision, detecting discontinuities and edges often marks the initial stage of processing a scene, guiding the recognition of shape and form (Hubel and Wiesel, 1962; Marr et al., 1997). In contrast, auditory scene analysis lacks these direct boundaries and must identify and distinguish sounds without the spatial-temporal separation of sensory surface observed in vision.

Alternatively, auditory objects can be characterized by a combination of features such as pitch, timbre, and loudness (Bregman, 1990). However, the exact acoustic properties underpinning these perceptions, such as harmonic or temporal differences, are not immediately obvious to the listener (Schnupp et al., 2011). In other words, the process of defining and identifying auditory objects is inherently non-intuitive, since auditory waveforms unfold dynamically rather than exist as static entities. These signals often overlap, forming a complex hierarchy of multiple individual events that require constant monitoring. This dynamic nature presents unique challenges for the auditory system during scene analysis, as it must discern individual auditory objects in a seamless auditory landscape that is not separated by clear boundaries.

Moreover, auditory objects must be generalizable across different sensory experiences since sound features can change under different listening context or physical actions while the sources remain the same (King and Nelken, 2009). For instance, you can recognize your teacher's voice regardless of your seating position in the classroom, even though varying distances can affect loudness; or you can recognize the melody of 'white Christmas' regardless of the speed at which it is played. The distinct physical properties of auditory inputs result in unique neural mechanisms for resolving this generalization process (Ison and Quiroga, 2008). The following section will review the crucial strategies for unravelling how the auditory system resolve the issues of grouping auditory objects and generalize across identity-preserving changes.

1.1.3. Auditory Cues in Complex Acoustic Environments

Understanding how we perceive and segregate sounds in noisy environments is essential for comprehending the formation of auditory objects and addressing the challenges associated with auditory processing disorders. This complex process is found to be

influenced by various auditory cues and principles that help determine whether sounds are perceived as coming from the same source or from multiple distinct sources.

1.1.3.1. Integrative Processes in Auditory Perception

Auditory perception heavily relies on memory due to the temporal nature of sound. Unlike visual objects, which can be represented in their entirety almost simultaneously, auditory objects unfold over time and often overlap with other sound. This temporal characteristic requires the auditory system to retain information about the initial parts of a sound as it processes the latter parts. A wealth of evidence indicate that echoic memory is involved in facilitating the perception of extended sound sequences by preserving auditory information for a few seconds, aiding in pattern recognition and temporal integration (Snyder and Alain, 2007). However, rather than directly storing raw sensory inputs into memory, the first step for the auditory system is to compress those data into a more condensed form for memory storage and further processing. This process is hypothesized to be primarily driven by two integrative approaches: sequential and simultaneous integration.

1.1.3.2. Sequential Integration

Sequential integration, supported by memory, is vital for connecting sounds that unfold over time. This process allows for the interpretation of patterns embedded within temporal sequences, making it essential for recognizing signals such as speech and music. It permits the brain to construct a continuous auditory stream from separate acoustic events over time and to extract various types of information (i.e. acoustic features, adjacent/non-adjacent sound dependencies, the rhythm in music, and the meaning of speech) across different time scales. For instance, the ability to recognize patterns in auditory sequences emitted by the same source, such as the distinct rhythm of a friend's footsteps can be explained by this approach (Yabe et al., 2001; McDermott et al., 2013; Baumgarten et al., 2021). In music, the brain interprets each note not in isolation but in relation to its rhythmic and melodic context. This is supported by that it is simple to recognize Beethoven's Moonlight Sonata regardless of the pianist, the piano used, or variations in tempo—whether it is played rapidly or slowly (Vuust et al., 2022). Sequential integration is crucial for speech perception, as it enables the distinction of phonetic elements and the construction of words and phrases. Even in noisy environments or when parts of speech are obscured, this process allows listeners to perceive complete words rather than disjointed groups of phonemes (Repp, 1988). Another key functionality of sequential integration is its ability to synthesise the constant properties of an auditory signal over time. Even though the physical properties of the sound may change as the source's action varies (for example, a lecturer's voice may change in intonation or pauses, or loudness can change when they move to different position), you still maintain a stable perceptual representation of your lecturer’s voice.

1.1.3.3. Simultaneous Integration:

Serving as a complement to sequential integration, simultaneous integration allows interpretation and organisation of sounds from various frequency regions, seamlessly merging them into a single auditory object. This integration is useful in complex auditory environments where multiple sources emit sounds at the same time. For instance, during a symphony, where numerous instruments play at once, we perceive their collective output as a harmonious whole, not as isolated instruments. This distinctive skill of the auditory system allows it to consolidate multiple sound sources into one coherent auditory stream. The ability to meld these inputs into a unified auditory perception is particularly important for appreciating music. However, this integration also implies that the auditory system may lack the sensitivity to distinguish individual sources as effectively as other sensory systems, such as vision. A likely reason for this is that auditory system has been evolved for early 'warning': to rapidly identify the potential nature of sound sources that have not yet been detected by other modalities—whether those are living beings? And what immediate actions might be necessary in response. (Bregman, 1990; Ragert et al., 2014; Winkler and Denham, 2024)

1.1.3.4. Essential Cues and Strategies for Auditory Object Formation

As discussed above, auditory processing demands sophisticated temporal analysis capabilities because sounds often arrive as mixtures rather than isolated elements. The auditory system's ability to parse these overlapping sounds into distinct streams or objects is crucial to understand complex auditory environments. The brain must track and integrate these time series - a process that is more temporally demanding than the relatively static and spatially ordered visual processing. Various cues and strategies have been hypothesized to play an important role in auditory grouping (Bizley and Cohen, 2013).

The Gestalt principle, developed by German psychologists in the early 20th century, describes how perceptual systems group sensory elements into a coherent whole. For example, densely occurring sounds, such as rhythmic patterns in music phase or syllables in words, are perceived as associated because the timing of the sounds helps to form unique linguistic units or musical phases (Ravignani et al., 2019). Similarly, when sounds have the same characteristics, such as pitch, timbre, or volume - such as the different notes played by a cello in a symphony orchestra - those sounds are heard as a cohesive entity, distinguished from other groups of instruments. In addition, the principle of closure in the auditory experience, that is, the mental completion of missing elements in the sound sequence by the brain, allows continuous perception even when the sound is intermittent. This phenomenon indirectly reflects the brain's ability to automatically group certain information into a coherent entity (McWalter and McDermott, 2019). In addition, auditory elements that are related in time and space provide important clues for the auditory system

to bring these elements together, as demonstrated by the unison performance of a choir or instrumental ensemble. Overall, those strategies suggest the auditory brain is capable of utilizing specific universal principles to construct a stable internal representation of the ever changing environment.

1.1.4. Modelling Auditory Scene Using Pure Tones

Nevertheless, the mechanisms through which the auditory system processes sound are complex and multifaceted, posing significant challenges for researchers. To overcome these difficulties and gain a clearer understanding of basic auditory processing in a controlled and replicable environment, accumulated research have employed pure tones to model the auditory scene (Fletcher, 1940). These controlled auditory stimuli allow precise manipulation of sound properties, including frequency, intensity, and duration, which is critical to isolate specific auditory mechanisms and thoroughly understanding how these temporal and spectral features affect perception and cognition.

In particular, tone-pips, especially those lasting 50 ms, which match the latency of phonemes in speech, are commonly used in many auditory research subfields. This approach was inspired by pioneering research such as that carried out by Eimas et al in 1971. In their study, the researchers explored how infants process language through an approach that combines synthetic speech sounds and high amplitude sucking response techniques, examining their ability to recognize phonemic distinctions. Infants were exposed to pairs of synthetic speech sounds distinguished by sound onset time - a key acoustic cue for phonemic differentiation. The results showed that even one-month-old infants showed a clear preference for sounds that crossed phonemic boundaries and exhibited a more pronounced response when those boundaries were crossed. This suggests that humans have an innate ability to integrate and classify short sound durations, similar to individual phonemes. (Eimas et al., 1971)

Further building on this, the study by Di Liberto et al. (2015) investigated the neural basis of phoneme-level speech processing using EEG. This research found that low-frequency cortical entrainment is sensitive not only to the acoustic properties of speech but also to its phonemic classification. By employing linear regression models, the authors linked EEG responses to continuous natural speech and its time-reversed version with various speech representations, including phonemic and phonetic features. They found that models which combined acoustic detail and phonetic categorisation were the most accurate at predicting EEG data, revealing that the brain response measured by EEG reflects more than passive acoustic tracking; it also mirrors higher-level speech-specific processing. Notably, the study also found that the model's predictive accuracy was reduced when time-reversed

speech was analysed, suggesting the significance of intelligible speech in eliciting distinct EEG responses. This research further reinforces the brain's capability to incorporate phonemic structures, advocating the use of tone-pip sequences to model the auditory scene. (Di Liberto et al., 2015) Importantly, these findings are further supported by recent empirical research. This research, which utilized intracranial recordings, has robustly demonstrated that the integration window in the auditory cortex can be as brief as 30 milliseconds (Norman-Haignere et al., 2022).

In this thesis, I will utilize pure tone-pips to model auditory patterns, focusing on investigating the neural mechanisms that enable the brain's profound ability to discern and represent these patterns. Specifically, the majority of the tone durations employed across most studies are set at 50 ms. However, in the first study, some tones are generated with a shorter duration of 25 ms (more details are provided in chapter 2). To guide the hypothesis formulation and facilitate the interpretation of results, this thesis will extensively utilize predictive coding theory. The following section will provide a comprehensive review of this theoretical framework.

1.2. Predictive Coding Theory

Predictive coding was initially proposed by philosopher Hermann von Helmholtz, who believes that perception is akin to unconscious inference, that is, making predictions about the world based on previous experience (Turner, 1977). This foundational idea has evolved significantly, now positioning predictive coding as a central framework for understanding brain function across perception (Kersten et al., 2004), cognition (Caucheteux et al., 2023), and motor control (Press et al., 2011). Unlike the traditional view of the brain as merely responding to stimuli, predictive coding portrays it as an active predictive machine that constantly creates and refines internal models of the world. This dynamic model building is driven by the brain's ability to generate predictions about sensory information based on past experience and the current environment (Rao and Ballard, 1999; Friston, 2005; Huang and Rao, 2011; Bastos et al., 2012).

Specifically, the brain employs these internal models to anticipate sensory inputs, resulting in prediction errors when the actual inputs deviate from expectations (Friston, 2005). These errors are crucial for refining the models and guiding the brain to adjust its predictions to better match the incoming data. The components involved in this adjustment process are called precision (Friston, 2010; Yon and Frith, 2021), which represents the confidence or inverse variance of the internal model. Suppose it moderates the effect of prediction errors (Friston, 2010).

Predictive coding operates through hierarchical processing, where higher cognitive functions can influence lower-level sensory perceptions, and vice versa. This interaction occurs through a combination of feedback and feedforward loops that integrate higher cognitive expectations with incoming sensory data (Lee and Mumford, 2003; Friston, 2008; Shipp, 2016). Accumulating computational modelling work (Knill and Pouget, 2004; Tenenbaum et al., 2006; Daunizeau et al., 2010; Skerritt-Davis and Elhilali, 2021) and experimental observations (Battaglia et al., 2003; Knill and Saunders, 2003; Barascud et al., 2016; Skerritt-Davis and Elhilali, 2021) show that the underlying mechanism of this process is primarily based on the principles of Bayesian reasoning. Specifically, the brain constantly calculates the probabilities of different outcomes, using previous knowledge (predictions) and new sensory evidence (actual input). Although review from Aitchison and Lengyel, (2017) suggests that various computational models, not just Bayesian principles, may underlie these brain calculations (Aitchison and Lengyel, 2017), the main goals remain consistent: These computational processes allow the brain to constantly update its beliefs to optimize perception and action in a changing environment.

1.2.1. Bayesian Processing in Perception

Predictive coding can be seen as a specific application of the Bayesian principle, where Bayesian processing systematically simulates how the brain integrates prior knowledge with incoming sensory data. This integration allows the brain to make informed inferences about the world, thus constantly updating its internal representations and facilitating the learning of new knowledge. This approach, based on Bayesian probability theory, refines beliefs based on new evidence through a series of updates. Initially, the brain establishes a prior probability (P(H)) that represents the likelihood of a hypothesis before processing a new sensory input. This priori is formed based on past experience/memory or an innate expectation of what is likely to happen. For example, anticipating a phone call from a friend might be based on how often they call you and when they usually call (Knill and Pouget, 2004; Friston, 2012).

Theoretically, when the brain integrate this likelihood with what it already believe (the prior knowledge), a mathematical method called the Bayesian formula is used to update the belief. This formula helps the brain calculate the updated probability of its hypothesis after taking the new evidence into account. The formula is: $P(H \mid E) = \frac{P(E|H) \times P(H)}{P(E)}$. $P(H|E)$ is the updated probability of the hypothesis after considering the new evidence E. $P(E|H)$ is the likelihood of the evidence assuming the hypothesis is true. $P(H)$ is the brain's initial belief in the hypothesis before receiving new sensory evidence. $P(E)$, the denominator, is the total probability of observing the evidence under all possible hypotheses, and it serves to

normalize the result. This normalization ensures that the probabilities for all hypotheses add up to one, making them valid probabilities. This formula allows the brain to revise its beliefs logically and systematically based on newly acquired information (Doya, 2007).

To contextualize this in an intuitive scenario, consider identifying the source of a sound in a home setting. For example, if you hear a beep, prior knowledge might suggest different hypothetical sources, such as a household appliance, a person, or an external noise, with initial hypothesis probabilities of 0.6, 0.3, and 0.1, respectively. Once you hear a beep again, and this provides you with more evidence that the beep pattern is more consistent with the microwave hypothesis (probability 0.8 if true), this particular likelihood, combined with a lower probability that it came from a person (0.1) or an external one (0.05), is used to update the belief about the source of the beep. Adding these weighted possibilities gives the total evidential probability $P(E)$, which totals 0.515 in this example. Then, the posterior probability that the sound came from a household appliance rose sharply and was calculated to be about 0.93, indicating a strong belief that the microwave was the source. Thus, this Bayesian framework not only helps to interpret everyday auditory scenes by balancing prior expectations with new sensory data, but also highlights the dynamic and probabilistic nature of perception, constantly adapting to incoming information flows to optimize understanding and interaction with the environment.

Step-by-Step Explanation of the Calculation:

This is the total probability of 'hearing the beep' given those 'hypothetical sources', using the formula:

$P(E) = (P(E|H_{appliance}) \times P(H_{appliance})) + (P(E|H_{people}) \times P(H_{people})) + (P(E|H_{external}) \times P(H_{external}))$

Plugging in the numbers:

$P(E) = (0.8 \times 0.60) + (0.1 \times 0.30) + (0.05 \times 0.10) = 0.48 + 0.03 + 0.005 = 0.515$

Calculation of Posterior Probability (P(H|E))

For the probability of appliance hypothesis:

$P(H_{appliance}|E) = P(E|H_{appliance}) \times P(H_{appliance}) / P(E) = (0.8 \times 0.6) / 0.515 \approx 0.93$

The McGurk effect is a classic example that illustrates how our perceptions are shaped by combining prior knowledge with new sensory information, a process well explained through Bayesian inference. Typically, we hold strong beliefs about how certain lip movements correspond to specific phonetic sounds—for example, expecting a "ba" sound

when lips close together. However, the McGurk effect presents a scenario where the brain encounters conflicting evidence: while the auditory input suggests "ba-ba", the visual input clearly shows lip movements that correspond to "ga-ga". In terms of Bayesian principles, the perceptual system recalculates the probabilities to resolve this sensory discrepancy. The likelihood of observing "ga-ga" lip movements aligned with a "ba-ba" auditory signal is less likely in terms of past experience. Consequently, the brain updates its beliefs, integrating the strong prior expectations with the new, conflicting sensory inputs. This integrative process typically results in a new, compromised perception, such as "da-da," a sound that was neither heard nor visually indicated but emerges as a coherent synthesis of both auditory and visual data (Green et al., 1991). Other examples such as speech perception in noisy environment (Macleod and Summerfield, 1987) and adaptation to altered auditory feedback (Houde and Jordan, 2002), further support the involvement of Bayesian processing in our perceptual modalities.

1.2.2. Hierarchical Models of Predictive Coding

Predictive coding is fundamentally based on the brain's hierarchical structure, wherein higher level (i.e. cognitive) functions influence at lower levels such as sensory processing and vice versa, enabling a dynamic interplay that continuously refines our perception based on newly sampled information. This structure is backed by electrophysiological and anatomical evidence showing that multiple brain regions, such as the auditory (Jasmin et al., 2019; Norman-Haignere et al., 2022), visual cortex (Rao and Ballard, 1999), and pre-frontal cortex (Badre and D'Esposito, 2007) are hierarchically organised.

From the computational perspective, opinions from Friston (Friston, 2002, 2005), Mumford and colleagues (Mumford, 1994; Lee and Mumford, 2003) illuminate the hierarchical organization while combines predictive coding and Bayesian inference. This model posits the neural system as an active hypothesis-testing machine. According to the model, high-level brain areas generate predictions based on learned contextual information, which are then sent to lower-level processing units to test on current inputs. When a mismatch ('prediction error') between expected and actual sensory signal occurs, the brain works to resolve this discrepancy by repeatedly adjusting how its different levels communicate over the presentations of the repeating stimulus. This involves fine-tuning the connection strengths within its hierarchical structure, allowing each level to update its expectations based on new information. These modifications result in the reduced future prediction errors, enhancing its ability to accurately interpret sensory inputs (This phenomenal is manifested as repetition suppression (Summerfield and de Lange, 2014) or expectation suppression (Todorovic and Lange, 2012)). When the prediction aligns with the

sensory input, the 'minimized' prediction error leads to an optimal Bayesian estimate of the sensory input. This allows the brain to choose the 'best guess' model. In this hierarchical structure, each layer, from lower to higher, refines the brain's predictive accuracy by systematically processing and integrating sensory information. Lower layers handle more detailed or raw sensory data, progressively abstracting and refining this information before passing it to higher layers. Higher layers, capable of representing more complex and contextual knowledge, refine these initial predictions, adjusting them based on broader understandings of the external world. This layer-by-layer enhancement ensures that each stage contributes to increasingly accurate predictions about environmental scene.

From the neural viewpoint, hierarchical processing is observed across various sensory modalities. In vision, the process begins in the retina, which captures and initially processes visual stimuli before passing them up through the brain's hierarchy to structures like the lateral geniculate nucleus (LGN) and finally to the visual cortex. Each level refines these initial predictions by incorporating more contextual information and feedback from higher levels, continuously working to reduce prediction errors (Huang and Rao, 2011).

Similar insights are also provided in auditory research. The influential work by Wacongne and colleagues specifically illustrates the use of hierarchical predictive coding in auditory perception. They introduced a paradigm which involves two categories of sounds, each created from distinct sets of superimposed sine waves. These sounds were presented in three configurations: standard sequences, which repeated the same tone five times to establish a predictable pattern; deviant sequences, where the pattern was broken by an unexpected fifth tone following four repeated ones; and omission sequences, in which the expected fifth tone was left out, deviating from the anticipated five-tone sequence. Their MEG and EEG results showed significant mismatch negativity responses when auditory sequences deviated unexpectedly from a pattern, with especially pronounced responses to omissions—where a predicted tone was absent. This indicates that the brain not only anticipates expected auditory patterns but also effectively adjusts its expectations based on the presence or absence of expected stimuli, illustrating a complex, multi-level processing mechanism for prediction and error correction across various cortical areas (Wacongne et al., 2011).

In addition to basic auditory stimuli, another empirical research by Caucheteux et al. (2023) reported similar results in speech processing using fMRI. This study analysed brain activity while participants listened to recorded stories, which revealed that the brain formulates predictions that span from immediate next word sounds to long-range, contextual linguistic constructs. The findings highlighted a multi-level predictive system where the frontoparietal cortices forecast complex narrative elements such as syntax structures, extending beyond the simpler phoneme processing typically managed by the temporal cortices. Moreover, the study revealed that higher cognitive regions, such as the prefrontal

cortex, engage in more advanced predictive tasks. These areas handle broader contextual predictions and manage information over longer timescales, which suggested a sophisticated and hierarchical approach to processing and anticipating linguistic information (Caucheteux et al., 2023).

More importantly, and above findings come to support the opinion that the hierarchy of predictive coding in the brain correlates with the brain's intrinsic time scales (Kiebel et al., 2008), on which different cognitive functions operate over specific durations. Higher cognitive functions, such as those managed by the prefrontal cortex, are engaged over longer periods to handle complex, anticipatory tasks. In contrast, lower-level sensory areas rapidly process immediate and detailed sensory stimuli. This distinction, both structural and functional, allows the brain to allocate neural resources more efficiently and minimize redundancy in processing information.

1.2.3. Neural Evidence of Prediction Error

Prediction error is the fundamental concept in predictive coding theory, it refers to the discrepancy between expected sensory input (top-down predictions) and actual sensory input (bottom-up data). The brain uses this error to refine its higher-level beliefs, adjusting its internal models to better match reality. This adjustment is achieved through feedback mechanisms that send error signals back up the neural hierarchy. As a result, the feedback mechanism enhances the accuracy of future predictions and optimises the brain's internal model.

Neural and physiological evidence of prediction error have been substantiated through various methodologies including brain imaging, electrophysiology, and experimental psychology (Bastos et al., 2012; Kok and de Lange, 2015; Kok, 2016; Shipp, 2016; Heilbron and Chait, 2018; Tabas and Kriegstein, 2023). For instance, the study by Arnal et al. (2011) investigated neural processing of prediction errors in audio-visual speech perception by exploiting the inherent delay between visual and auditory speech signals to create congruent and incongruent conditions. Using MEG, the research revealed distinct patterns of neural oscillations—slow delta oscillations in higher-order speech areas like the superior temporal sulcus under congruent conditions and shifts to low-beta and high-gamma oscillations in multisensory areas during incongruence. These findings underscore the brain's utilization of specific oscillatory dynamics to code for prediction errors (Arnal et al., 2011).

At the neural anatomy level, generation of prediction error is thought to be a fundamental brain function, which is encoded across a wide network of brain regions such as cortical and subcortical areas (Den Ouden et al., 2012). These include the sensory cortices (visual (Rao and Ballard, 1999), auditory (Tabas and Kriegstein, 2023), and

somatosensory (Yu et al., 2022)), which adjust neural responses based on discrepancies between expected and actual sensory data. The premotor and motor cortices also function in this way, especially in aligning motor actions with expected outcomes (Shadmehr and Krakauer, 2008). Furthermore, the frontal cortex, which is associated with higher level functions, has been found to be sensitive to unpredictable sensory deviations (Näätänen et al., 2005; May and Tiitinen, 2010; Dürschmid et al., 2016). Evidence from subcortical areas, such as those supported by Iglesias et al. (2019), highlights the hierarchical processing of prediction errors. Specifically, the study utilized fMRI to show how the ventral tegmental area and substantia nigra specifically process low-level prediction errors related to direct sensory outcomes, whereas the basal forebrain manages higher-level errors associated with stimulus context-outcome contingencies (Iglesias et al., 2019).

1.2.3.1. Neural Transmitters that Modulate Prediction Error

At the cellular level, insights from pharmacological research suggests that various types of neural transmitters are associated with prediction error processing. For example, the study by Marshall et al. (2016) propose that Noradrenaline (NA), particularly sourced from the locus coeruleus (LC), is associated with rapid updates in perceptual belief about the volatility of the environment, facilitating the brain's response to unexpected environmental change. This response to volatility helps maintain focus on relevant stimuli, thereby enhancing adaptability to new information. Pharmacological manipulation and behavioural task results supports this, which suggested that blocking noradrenaline receptors can change the rate at which beliefs about environmental volatility are updated (Marshall et al., 2016). Similar insights from the modelling work (Sales et al., 2019) also pointed out the critical role of the LC-NA system in augmenting cognitive flexibility within dynamic environments, responding to prediction errors with suitable adjustments in learning rates and belief updates.

In addition to the NA, Marshall et al. (2016) also examined the role of acetylcholine (ACh) in coding uncertainty within dynamic environments, a process relevant to prediction error. Their findings suggest that when ACh receptors are blocked, participants' ability to quickly adapt to changes significantly declines. They interpreted that ACh plays an essential role in attributing uncertainty, whether it is due to fluctuations within a stable environmental context defined by probabilistic associations, or to larger environmental changes following a contextual shift (Marshall et al., 2016).

The study by Iglesias et al. (2021) further provides evidence for the mechanism of ACh in modulating brain responses to prediction errors. The researchers used ACh blockers (Biperiden) and enhancers (Galantamine) on human participants and measured the brain activity using fMRI. The experiment involved an audio-visual associative learning task to

explore the effects of ACh on low level and high level prediction errors. Participants were tasked with learning the predictive strengths of auditory cues (high or low tones) to determine which of two visual targets (a face or a house) would appear. This task was designed to include various levels of cue-outcome association strength (probabilities), creating a dynamic environment of volatility in which these long-term associations would change over time. The fMRI data and general linear models reveal that ACh significantly influences brain activity related to these errors. Using ACh blockers heightens the brain's response to low-level prediction errors, such as direct auditory cues and visual outcome discrepancies. However, it diminishes responses to high-level prediction errors, like changes in environmental volatility, specifically in the brainstem regions. Interestingly, this study also reported the observation of increased low level error response while use ACh enhancers, which appears to be controversial to the effects of blockers. These findings suggest a complex mechanism of acetylcholine in modulating prediction error in the hierarchical levels of the brain (Iglesias et al., 2021).

Dopamine, a crucial neurotransmitter involved in associative learning, is also closely linked to concept of prediction errors. It primarily signals positive prediction errors, i.e. unexpected rewards (Nasser et al., 2017; Lerner et al., 2021). As demonstrated by Takahashi and colleagues, dopamine neurons also encode errors in predicting sensory aspects of expected rewards. Their research shows that dopamine neurons respond to discrepancies not only in predicted and received rewards, but also in the sensory features associated with these rewards (Takahashi et al., 2017). This implies that dopamine signals may assist the brain in adjusting not only to unexpected rewards, but also to unforeseen environmental changes related to those rewards. It indicates a probable neural mechanism that integrate sensory information into the reward prediction framework. This was further supported by Iglesias et al. (2021). Using similar pharmacological manipulation methods as those used with ACh, they revealed that dopamine is associated with immediate, sensory-related (low-level) prediction errors. Importantly, the study also reported the complex interactions of dopaminergic and cholinergic systems, suggesting that these chemicals do not work independently but cooperate in a more sophisticated mechanism (Iglesias et al., 2021).

1.2.3.2. Neural Correlates of Prediction Error

From the perspective of neurophysiological recordings, Mismatch negativity (MMN) is closely associated with concept of prediction errors—its response pattern reveals the discrepancy between the brain's predictions based on its established environmental model, and the actual sensory inputs. MMN is a crucial component of the event-related potential (ERP) in brain activity, observed when an unexpected stimulus deviates from a repetitive

pattern of stimuli, even without conscious attention (Näätänen et al., 2005). The response typically occurs between 100 and 250 milliseconds after the presentation of a deviant stimulus. It is primarily found to be associated with the auditory and frontal cortices, serving as a rapid indicator of sensory surprise.

Notably, the review from Garrido and colleagues describe MMN as a reflection of Bayesian inference processes, where deviations from predicted sensory inputs lead to prediction errors that prompt the brain to adjust its predictions, with the MMN magnitude reflecting the degree of surprise (Garrido et al., 2009). Wacongne et al. (2012) provide computational insights on this by proposing a neuronal model suggesting that MMN arises from synaptic adjustments made in response to hierarchical prediction errors, thus enhancing the brain's predictive accuracy (Wacongne et al., 2012). Baldeweg (2006) reviewed how the auditory system processes repeated sounds via predictive coding, noting that expected sounds typically produce reduced neural responses unless an unexpected 'anomaly' occurs, thereby eliciting MMN (Baldeweg, 2006). This phenomenon demonstrates the brain’s mechanism for minimizing prediction errors. Extending this principle, Stefanics and colleagues apply the concept of MMN in visual domains, suggesting that MMN's role in addressing prediction errors has cross-modal applications, further emphasizing this neural signature as the reflection of perceptual inference and learning across multiple sensory modalities (Stefanics et al., 2014).

Apart from MMN, Repetition Suppression (RS) and Expectation Suppression (ES) are two phenomena closely associated with the concept of prediction error, which are both frequently explored in perception literature (Todorovic and Lange, 2012; Barbosa and Kouider, 2018; Tang et al., 2018). RS manifests as a decrease in neural activity that follows the repeated presentation of the same stimulus. Physiological studies suggested that it is the reflection of neural habituation or sensory adaptation (Summerfield et al., 2008; Nelken, 2014). This reduction generally occurs without regard to context, mainly driven by changes in neural fatigue or synaptic efficiency within sensory-specific brain areas (Thompson and Spencer, 1966). On the other hand, ES is related to the brain's anticipatory mechanisms; when a stimulus is expected, the resulting neural activity is lessened due to reduced novelty or surprise (Han et al., 2019). This makes ES highly context-dependent, and requires higher level processes such as memory and attention to manage expectations (Todorovic and Lange, 2012; Kaposvari et al., 2018; Feuerriegel et al., 2021). Despite their unique underlying mechanisms, RS, which is rooted in sensory exposure, and ES, which is based on contextual prediction, both reflect the brain's hierarchical approach to sensory processing, in line with predictive coding.

In specific, prediction error minimisation represents accurate expectations, while inaccurate predictions generate larger errors that facilitates adjustments to the brain's internal predictive models. ES occurs when the brain's predictions are accurate that the

incoming stimulus matches these predictions, resulting in minimal prediction errors and thus reduced neural activity. Similarly, RS arises when a stimulus is repeatedly presented, and the sensory units become less responsive due to the lack of novelty. This leads to a reduction in neural activity as the prediction error decreases with each repetition. Therefore, within the framework of the hierarchical model of predictive coding theory, both RS and ES can be seen as the indirect manifestations of prediction error (Mayrhauser et al., 2014). However, some studies challenge that RS is not directly linked to the concept of prediction error since no neural evidence has been measured in visual cortex (Solomon et al., 2021).

Nonetheless, the above models present interpretations to the process from which the brain tests and refines its hypotheses across different layers, thus improving the efficiency of sensory processing.

1.2.4. What is Precision

How does the brain update its internal model and modulate prediction errors? The concept of precision, as initially proposed by theorist Karl Friston within the framework of predictive coding, is central to this adaptive process (Friston, 2005, 2010). Precision is defined as the confidence of inferred reliability of sensory inputs, which weighs prediction errors within hierarchical brain models. Their amplitude is usually represented as inverse variance of inferred predictive distribution. Physiologically, precision is hypothesized to modulate synaptic gain on prediction error units (Friston, 2010; Yon and Frith, 2021). The principle proposed that by optimizing precision, the brain can effectively manage how much influence different sensory inputs or prediction errors have on its overall perception/cognition and neural response processes. Such regulation allows the brain to allocate more resources to deal with unforeseen environmental uncertainty.

The neural mechanisms underlying precision remain largely elusive. Evidence from neural transmitter studies suggest acetylcholine (ACh) and noradrenaline (NE) are crucial neuromodulators in this process. Particularly, tonic Ach has been suggested to be correlated with contextual uncertainties (Yu and Dayan, 2005). Furthermore, a recent study on rats demonstrates how ACh significantly enhances the precision of neural responses to sensory inputs by sharpening responses to auditory discrepancies, thus optimizing sensory processing and adaptation (Pérez-González et al., 2024). In addition to ACh, NE functions as altering the synaptic efficacy, enhancing strong prediction error signals, and suppressing weaker ones. This effectively allocates the brain's resources on the most unexpected or uncertain inputs (Shipp, 2016). Ferreira-Santos, (2016) indicates that this modulation aligns with the GANE (glutamate amplifies noradrenergic effects) model, which posits that NE's primary function is to adjust the precision, or confidence, applied to prediction errors across

sensory modalities, enhancing the signal-to-noise ratio in neural representations (Ferreira-Santos, 2016).

From the neural basis standpoint, emerging evidence have pointed out that precision might be related to inhibitory mechanisms (Natan et al., 2017; Schulz et al., 2021; Richter and Gjorgjieva, 2022; Yarden et al., 2022). For example, Bastos and colleagues discussed the intrinsic connections among excitatory and inhibitory populations within cortical columns, particularly those involving inhibitory neurons, crucially contribute to the regulation of neural activity. These inhibitory neurons, located primarily in the granular and infragranular layers, are key to controlling the gain and precision of neural responses, fine-tuning the processing of prediction errors. The group suggested that this regulation allows for the dynamic adjustment of neural responses, and help scales the responses appropriately to the uncertainty or reliability of sensory inputs. Supported by key evidence from cortical studies, the neural mechanisms help the brain prevent from overloading predictable inputs while enhancing responsiveness to novel stimuli (Bastos et al., 2012).

Similarly, Shipp discussed on how inhibitory neurons within cortical layers rich in interneurons modulate the precision of predictions. By adjusting the gain of pyramidal cells involved in prediction error signalling, these inhibitory neurons optimize the brain's response based on sensory inputs' predictability and relevance (Shipp, 2016). Furthermore, one recent study employed the dynamic causal modelling to investigate how inhibitory mechanisms modulate precision within predictive coding. It highlights that precision is dynamically regulated by the inhibitory processes, especially for superficial pyramidal (SP) cells. By integrating EEG and MEG data, the study reveals that precision weighting is closely tied to the self-inhibition of SP cells, and inhibitory neurons are crucial in modulating those cells' gain. This modulation significantly influences how prediction errors are processed and alters the neural response to surprises, depending on their predictability (Lecaignard et al., 2022).

In summary, precision within the predictive coding framework is a crucial concept that underpins the brain's ability to integrate and assess sensory information and prediction errors. Efficient regulation of precision is not only crucial for normal sensory processing but also for maintaining brain health. Disruptions in the normal functioning of precision are linked to various neuropsychological disorders, such as schizophrenia and autism, where it impacts sensory integration and cognitive responses (Van de Cruys et al., 2014; Sterzer et al., 2018; Lecaignard et al., 2022). To deepen our understanding of this concept, this PhD thesis utilised auditory modality to address some of the questions underpinning this process (see Chapter 3 and 5). The next section will focus on reviewing literature related to the main topic of this thesis: the brain's sensitivity to patterns and the emerging understandings about its neural mechanism.

1.3. Sensitivity to Patterns

1.3.1. Statistical Learning in Auditory Scene

The natural world is governed by physical laws that lead to predictable sound patterns and structures. For example, the sound of rain falling often exhibits regular rhythmic patterns. This regularity comes from the principles of physics: formation of raindrops and speed falling from sky. The size and frequency of raindrops dictates the tempo and intensity of the sound pattern. Similar for the harmonic series in musical instruments. When sources such as violins or guitars produce sound, the vibrations of the strings generate a fundamental frequency along with a series of harmonics. These harmonics are integer multiples of the fundamental frequency, creating a predictable and regular harmonic series that is a direct result of the physical properties of the strings.

In neuroscience, "statistics" refers to the underlying patterns and structures within data that the brain learns to recognise, and the purpose is to make predictions and interact with the environment. The neural representations of the predictable patterns, or 'regularities', extracted from sensory inputs, are believed to serve as the source action inferences and environmental navigation (Winkler et al., 2009; Winkler and Denham, 2024). From technical perspective, these statistics are the distributions, relationships, and regularities across sensory inputs, in which an organism is exposed to over time.

Statistical learning refers to the brain's inherent capability to learn the rules in the sensory environment (Li et al., 2004; Berkes et al., 2011). For example, common rules that have been found includes probability distributions, which involves evaluating the frequency and likelihood of different sensory events (Winkler, 2003; Romberg and Saffran, 2010). It also involves analysing temporal and spatial correlations (Fiser and Aslin, 2002; Schapiro and Turk-Browne, 2015), examining the sequence of events or their relation to each other over time and space. Furthermore, the abilities to learn transition probabilities of sensory events (Schapiro and Turk-Browne, 2015), or even more complex network structures (Ren et al., 2022; Benjamin et al., 2024) have been demonstrated extensively. Nevertheless, regardless of what types of 'statistics' the brain learns, the ultimate purpose is to predict future events and minimize the response time.

The seminal research by Saffran and colleagues marked a pivotal point in our understanding of infants' capabilities for statistical learning by demonstrating that 8-month-

old infants can segment words from continuous speech streams. The authors achieved this without the explicit instruction or the reliance on acoustic cues such as pauses or intonation, but instead only the transitional probabilities between syllables were used as statistical cues. In the experiments, infants were exposed to synthesized, monotone speech streams comprising four, three-syllable nonsense words, repeated randomly without pauses. The primary cues for identifying word boundaries were the higher transitional probabilities within words, as opposed to those between them ('nonwords' or 'part-words'). The results demonstrated that infants could accurately distinguish between 'words' they had previously encountered and 'nonwords' or 'part-words'—sequences either not introduced during familiarization or that intermixed syllables across word boundaries. This finding highlights the infants' substantial reliance on the statistical properties of language to effectively differentiate linguistic stimuli (Saffran et al., 1996). Crucially, this innate ability to learn complex linguistic structures was not limited to linguistic stimuli but was also applicable to non-linguistic stimuli such as pure tones (Saffran et al., 1999; Gebhart et al., 2009).

Further extending the exploration of statistical learning across species, Hauser et al. (2001) investigated whether cotton-top tamarins could utilize this form of learning to process human speech patterns, using the same experimental paradigm as Saffran et al. (1996). The findings demonstrated that tamarins, akin to humans, could differentiate words from non-words and part-words based solely on the statistical characteristics of the syllable (Hauser et al., 2001). Additionally, Murphy et al. (2008) provided evidence that rats could learn and apply simple rules from sequences of stimuli to novel situations. In their experiments, rats were trained to associate specific three-element tone sequences, where the sequence structure followed a particular rule (such as XYX), with the reward of food. These sequences were created using stimuli that the rats had not been previously exposed to. The goal was to test whether the rats could generalize the rule learned from the training of novel auditory sequences, which either conformed or did not conform to the learned rule. Remarkably, the rats successfully learned that the sequences adhering to a specific structure were associated with food and were able to transfer this understanding to novel sequences composed of entirely new elements that followed the same statistical rule (Murphy et al., 2008). In exception of rats, other animals in nature, including Zebra Finches (Menyhart et al., 2015). Bengalese Finches (Takahasi et al., 2010), and Budgerigars (Spierings and Ten Cate, 2016) have also been found capable of learning statistics in auditory sequences.

1.3.2. The Neural Underpinnings of Statistical Learning

1.3.2.1. The Odd Ball Paradigm

Understanding the brain's mechanism of how it detect and respond to statistical structures in sensory input is an essential aspect for deciphering the brain. One of the primary experimental paradigms used to investigate these cognitive processes is the oddball paradigm. This paradigm typically involves presenting a sequence of repetitive standard stimuli interspersed with infrequent deviant stimuli, which differ in perceptual dimensions such as frequency, pitch, or orientation. Its versatility allows adaptation across various sensory modalities, notably auditory and visual systems, hence making it invaluable for examining neural mechanisms underlying perception, attention, and memory.

A seminal work by Näätänen and colleagues employed the oddball paradigm to explore auditory processing, leading to the discovery of MMN, an early ERP component triggered by deviations in auditory patterns. As discussed before, this component signifies the brain's automatic response to unexpected stimuli even without the top-down attention, suggesting the auditory system's high sensitivity to acoustic changes (Näätänen et al., 1978). Another relevant study by Escera et al., 2000, further demonstrated how the brain's involuntary attention is captivated by novel sounds using this paradigm, highlighting the distinct ERP components produced by 'oddball' sounds, particularly through the P3a and N1 components which is proposed to be reflecting involuntary attention shifts and early sensory processing (Escera et al., 2000).

The oddball paradigm is also extensively used in animal studies due to its simplicity. For instance, influential research by Ulanovsky et al, (2003) on cats revealed that neurons in the auditory cortex respond more robustly to rare, deviant tones than to common ones, a phenomenon known as stimulus-specific adaptation (SSA). The author suggested that this adaptation allows cortical neurons to distinguish between sound frequencies with exceptional precision—a trait referred to as "hyperacuity." Interestingly, their findings suggest that thalamic neurons did not exhibit similar probability-dependent changes, instead, SSA is primarily a cortical phenomenon, possibly linked to the neural mechanisms underlying MMN. (Ulanovsky et al., 2003)

The influence of voluntary attention on neural correlates of oddball paradigm are comprehensively investigated. For instance, Justen and Herbert (2018) focused on the spatio-temporal dynamics of auditory deviance and target detection. It implemented both passive and active auditory oddball paradigms, integrated with ERPs and Standardized Low-Resolution Brain Electromagnetic Tomography. Participants were exposed to standard and deviant tones (500 Hz vs. 1000 Hz) across passive and active listening conditions, revealing significant activations in several brain regions. For example, passive listening triggered

activations in the right superior temporal gyrus and bilaterally in the lingual gyri during the N1/MMN phases, and in the insulae during the P3 phase. Active listening, on the other hand, led to activations in the right inferior parietal lobule during the N1/MMN and across multiple cortical areas including the precuneus during the P3. The result demonstrated that different listening modes tend to engage distinct brain networks in processing the auditory deviance. (Justen and Herbert, 2018)

Collectively, the application of oddball paradigm offer profound insights for the neural mechanism of statistical learning. In addition to oddball paradigm, the local-global paradigm is another classic technique in this field, designed to study how the brain extracts hierarchical information by organizing stimuli into local elements which form part of a larger global structure. This approach allows researchers to examine how the brain integrate information across different hierarchical levels and has been instrumental in delineating the neural mechanisms underlying different layers of information processing.

1.3.2.2. Local Global Paradigm

David Navon pioneered this paradigm in 1977, utilizing it in the visual domain to investigate how people perceive hierarchical structures. In his seminal study, Navon presented visual stimuli composed of large letters (the global level) made up of smaller letters (the local level). He discovered that participants processed the global shape of the stimuli more rapidly and accurately than the local elements, a phenomenon termed "global precedence." This finding suggests that the human visual system is primed to recognize overall patterns before discerning finer details, implying a cognitive processing strategy where initial perceptions are shaped by general features, followed by subsequent attention and detail recognition. (Navon, 1977)

In the auditory domain, Koelsch et al. (2013) applied the local-global paradigm to explore how listeners process hierarchical syntactic structures in music, using J.S. Bach chorales. The study differentiated between regular hierarchical structures, where musical phrases adhered to expected tonal harmony, and irregular structures, where phrases were intentionally altered by transposing the first phrase down a fourth (e.g., from C major to G major) or up a major second (e.g., from C major to D major). These modifications disrupted the anticipated harmonic closure, creating the hierarchical irregularities while the local structure—individual melodic and harmonic integrity within phrases—remained unchanged. The brain responses revealed that the final chord from irregular structures prompted an early frontal negativity (150-300 ms), an initial detection of irregularities, followed by a later negativity (550-850 ms) formed by the deeper processing of the disrupted harmonic expectations. Interestingly, these findings were consistent across musicians and non-

musicians, which highlights a general feature of human auditory perception capable of processing hierarchical deviations in music. (Koelsch et al., 2013)

Additionally, a more comprehensive study by Maheu et al., (2019) applied the local-global paradigm to study brain responses to hierarchical sequences. This paradigm adjusts the complexity of auditory sequences by dividing them into five-tone chunks, each representing a unique pattern such as a series of repeated tones (e.g., XXXXX) or a pattern with one different tone (e.g., XXXXY). The study manipulated both within-chunk and between-chunk deviance by varying these patterns' occurrences within an experimental block, allowing the investigation of brain responses to local transition probabilities (within-chunk) and global pattern rarity (between-chunk) separately. Participants were instructed to focus on the sequences while their brain activity was monitored via MEG. The results suggest that early and mid-latency brain responses, such as the MMN around 250 ms, are responsive to item frequency and alternation. This indicates that the brain processes these factors over relatively short timescales. However, late brain responses, P300 as indicated by this study, reacted to the overall rarity of the patterns throughout the block, suggesting a longer integration timescale and illustrating a complex interaction between local details and global structural awareness in sequence processing. (Maheu et al., 2019)

1.3.2.3. Can Brain Learn Complex Statistics

As per the local and global paradigm, the brain's capability to respond differentially to various hierarchical levels of sensory information aligns with the hierarchical model of predictive coding. Aside from these traditional paradigms, increasing research is centred on directly evaluating the information or statistics within sensory inputs and identifying their associated neural correlates.

Skerritt-Davis and Elhilali (2018) suggests that natural sounds are complex and cannot be fully described using only basic statistics such as frequency or transition probabilities. Instead, fractals, which are patterns that display self-similarity across various scales, may be a better way to model the complexities in sound sequences. In this study, the authors utilized sequences of random fractal tone sequences, each defined by different entropy level, a metric of randomness or unpredictability. This allowed for a controlled stimuli of how entropy changes impact detection performance. In the experiment, participants were tasked with identifying changes in tone sequences, and the result revealed that the detection of entropy changes depended on both the magnitude of the change and the initial entropy level of the sequences. Brain responses recorded by EEG showed that larger deviations in adjacent frequency (ΔF), particularly in low-entropy contexts, resulted in stronger ERP responses. This suggests the brain is especially sensitive to unexpected changes in predictable (low entropy) auditory contexts (Skerritt-Davis and Elhilali, 2018).

Subsequently, the introduction of the Bayesian predictive inference model (D-REX) to simulate brain response aligns with the neural data, which further provided insights for the neural underpinnings of how the brain extract complex auditory statistics. The stimuli used in this study varied along two acoustic dimensions, using a random fractal structure to modulate entropy levels within auditory sequences. Participants in the EEG experiments were tasked with detecting changes in entropy-modulated sequences. They either listened to the nSP condition (sequences vary in spatial-pitch dimensions) or the nTP condition (sequences vary in timbre-pitch dimensions). The EEG results and model fitting suggest that local surprises in acoustic features linearly modulated the neural responses, while global, melody-level statistics induced a nonlinear integration across features. Specifically, the frontocentral network, visible in the initial processing stage with a latency window of 80-150ms post-stimulus, displayed neural activity closely aligned with local statistical changes in auditory features such as pitch and timbre. After this initial detection, the centroparietal network seemed to engage over a prolonged period to integrate these changes into a global context. The authors observed that this integration aligned with later ERP components within a time window of 300-800ms. Those findings support the hypothesis that the brain learns sensory information in the hierarchical manner and further suggest that the neural correlates for extracting statistics across multiple layers/time scales appear to be dissociable (Skerritt-Davis and Elhilali, 2021).

The natural environments we encounter daily are often stochastic and marked by uncertainty, making entropy an ideal quantification for modelling these variable conditions. However, in structured settings like language and music, elements often exhibit predictable relationships that entropy alone cannot fully capture. To address this, some research, such as the seminal work by Schapiro et al. (2013), introduced the concept of network structure to investigate the statistical learning in vision domain (Schapiro et al., 2013). This approach explored how discrete units are intricately interconnected within a larger framework, emphasising the dynamics of how individual elements influence one another.

The framework was recently tested by Benjamin et al. (2024) in auditory modality, in their experiment, participants passively listened to sequences of tones structured according to a "sparse community network," which comprised two clusters (communities) of tones that were densely connected (high transition probabilities) within and sparsely connected (low transition probabilities) between each other. The tones within or between the cluster were organised to maintain uniform transition probabilities. The task required participants to passively listen to these sequences, with a focus on the auditory input without making any active responses. MEG results showed rapid brain responses to changes within the sequences, occurring about 150 milliseconds after tone transitions, an indicative of a keen sensitivity to the network's structured relationships. Additionally, time-resolved decoding techniques revealed significant overlaps in neural representations of successive

tones, suggesting that the brain maintains activity from previous tones as new ones are perceived. This overlap suggests that the brain representation of tones are linked in a sequence, which is crucial for prediction and comprehension of contextual information. The study also found that neural responses were influenced by a novelty index from an associative learning model, and therefore the auditory processing is dynamically shaped not only by immediate spectral temporal changes but also by the accumulated structural knowledge. These findings highlight the brain's capacity for temporal integration and have profound implications for understanding the mechanisms of auditory memory, particularly how continuous auditory information is integrated and retained for tasks like monitoring speech and music.

1.3.3. Neural Correlates of Auditory Regularity Encoding

Researchers often come across this question: How does the brain retain accumulated knowledge and represent auditory context? This ability is widely demonstrated through the phenomenon of MMN, as reviewed previously. MMN serves as a critical neural mechanism, triggered when an auditory stimulus deviates from a repetitive pattern that the brain has learned. Such deviation prompts the MMN response, indicating that the brain has stored a representation of the pattern and recognizes deviations from it. Therefore, one hypothesis posits that MMN is indirectly indexing the memory trace of the established pattern. The MMN thus serves as an important marker for sensory memory. Alternatively, in terms of predictive coding theory, the MMN is hypothesized to arise from the brain's ability to create and maintain a model of auditory environment. This model helps the brain predict future events based on past experiences. When an incoming sound deviates from these expectations, the discrepancy triggers the MMN, which reflects the brain's detection of this violation of learned pattern.

One foundational study by Näätänen and colleagues first detailed how MMN can be elicited by deviations from a repetitive sequence of sounds, suggesting that the brain automatically generates a sensory representation of the auditory environment. This research laid the groundwork for proposing MMN as an index of sensory memory's role in detecting regularities (Näätänen et al., 1978). Building upon this, research by Winkler and colleagues (1996) demonstrated that MMN amplitude could be modulated by the predictability of the auditory sequence, directly linking MMN to the brain's expectation based on previous experiences (Winkler et al., 1996).

As MMN is an indirect and momentary marker of regularity tracking, is there a direct neural representation of regularity? The ongoing question of whether the neural correlates of regularity can be directly measured, was addressed in a comprehensive study by

Barascud et al. (2016). The authors conducted an in-depth investigation into auditory processing by using stimuli composed of 50-ms tone-pips organized into regularly repeating pattern (REG) and random (RAND) sequences. The study systematically manipulated the predictability of these REG patterns by varying the alphabet size (the number of different frequencies used within the sound pattern). Smaller alphabet sizes, which involve fewer frequencies repeated more frequently, resulted in more predictable sequences. Conversely, larger alphabet sizes introduced greater informational diversity, thereby reducing predictability. In contrast, random (RAND) sequences, defined as tone-pip sequences consist of randomly ordered frequencies, lacked deterministic pattern or regularity, serving as a baseline against which the structured REG sequences were compared. In the experiment, participants listened to those sequences while being instructed to pay their attention to a visual decoy task. The MEG recordings demonstrated a significant variation of the sequence evoked sustained response (DC) based on the predictability of the sequences. Specifically, REG patterns with smaller alphabet sizes elicited stronger sustained neural responses, indicating that the brain was more actively engaged when the auditory input was predictable. However, for the most unpredictable condition RAND, the brain exhibits lowest sustained response amplitude, compared with other predictable patterns.

Crucially, the dynamics of these sustained responses closely mirror the output pattern of Ideal observer model (IdyOM). This alignment suggests that the way the brain monitors information, as reflected by the sustained response dynamics, might be consistent with the statistical parameters tracked by IdyOM, which is based on the Prediction by Partial Matching (PPM) model employing a variable-order Markov model framework (Pearce, 2005). The fundamental principle of the PPM model is based on its assumption that the model has completed the initial integration processes and analyses the sequences symbol by symbol. In details, it creates predictive distributions for the next symbol through a synthesis of predictions from various sub-models, particularly n-gram models. An n-gram model uses sequences of 'n' adjacent symbols to generate conditional probabilities. For instance, to predict the next symbol following the sequence 'ABCAB', the model analyses occurrences of 6-grams like 'ABCABX', where 'X' is the variable component. The probability of 'C' being the next symbol, is determined by the frequency of 'ABCABC' relative to all 6-grams that begin with 'ABCAB'. The model output is quantified as the negative log probability of a tone occurrence at this position, conditioned on the portion of the sequence heard so far. Intuitively, this metric quantifies the 'surprise' experienced by the model when encountering each tone in a sequence. As the sequence become predictable, the information content typically drops, indicating that the tones are less surprising as the model adjusts to the emerging pattern (more details of the model will be provided in Chapter 2).

Despite IdyOM's resemblance of the brain responses, Barascud et al. (2016) discussed fundamental discrepancies between human auditory processing and the model.

Unlike IdyOM, which operates as if it possesses unlimited memory and computational resources, the human brain is constrained by neural and memory capacity, impacting its ability to relentlessly process and retain complex auditory sequences. This distinction is evidenced by the model's performance, where regardless of pattern complexity, IdyOM consistently reaches a uniform level of amplitude—quantified as information content—once a regular (REG) pattern is identified (See **Figure 1.1**). In contrast, in human listeners, the amplitude of sustained responses varies depending on the complexity of the pattern. This variation suggests that neural representations for sequence statistics in humans are significantly influenced by their computational capacities such as memory. Relevant questions were addressed in Study 1 (see Chapter 2), which aims to provide insights into these memory and information integration processes.

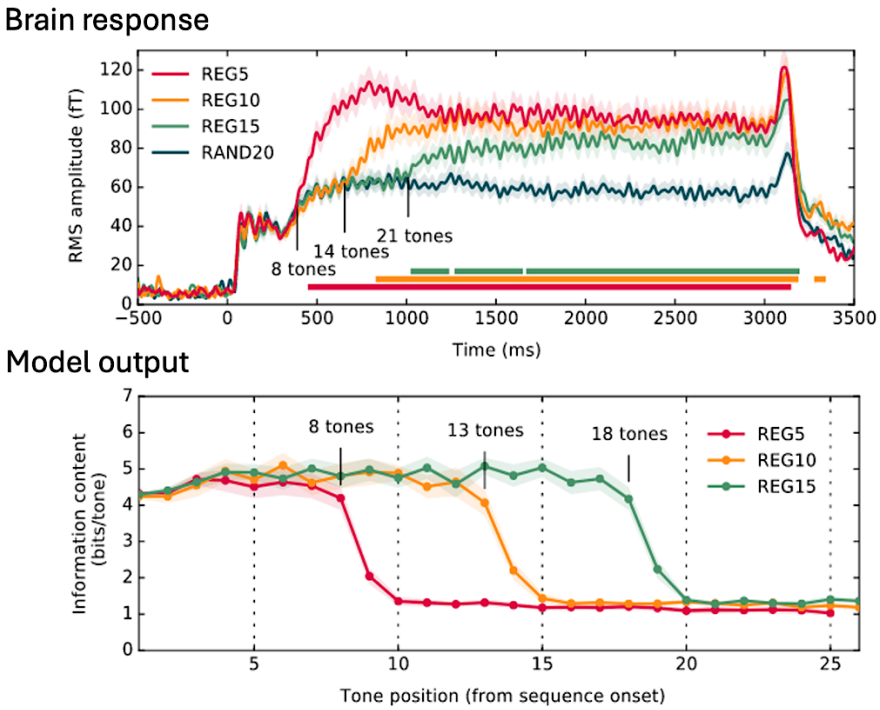

Figure 1.1. Brain response: The graph depicts the root mean square (RMS) of brain responses for pattern condition REG5, REG10, and REG15, as well as for the random sequence condition RAND20. The alphabet here represents the number of frequencies that make up the pattern or sequences. This representation covers the full stimulus epoch. The number of tones marked in each REG condition represents the point where the neural response of pattern starts to diverge from the control condition (RAND). Intervals indicated by lines below the brain response highlight the points where significant differences has identified between

each REG condition and RAND20. Model output: The model output displays the averaged information content (across trials) for each tone pip over the evolution of sequence processing for REG5, REG10, and REG15 conditions. The drop of information content indicate the point where the model discovered the pattern. (Barascud et al., 2016)

Back to the study, the authors propose that the increases in sustained response power are linked to the enhanced predictability in sequences for various patterned stimuli. Beyond the underpinnings by IdyOM, they suggest that this response seems to be associated with the inferred reliability or "precision" of sensory inputs—the key concept in predictive coding theory as reviewed previously. However, a significant limitation of their study is that IdyOM is not designed to track stimulus precision directly. Furthermore, the study exclusively uses either random or perfectly deterministic sequences. Such choice restricts the interpretation overall, as such deterministic patterns do not allow for subtle variations that could more effectively reveal how the brain adjusts its predictions under minor fluctuations in uncertainty or change. These conditions of slight variability are more common in natural environments and thus more relevant for understanding real-world sensory processing. To test this hypothesis, Zhao et al. (2024) introduced rapidly unfolding stochastic sound sequences, exploring how the human brain responds to these sequences, and whether the Bayesian prediction inference model (D-REX), which is designed to continuously monitor precision, could effectively benchmark neural dynamics or not. Their results demonstrated that the transitions between different stochastic tone patterns trigger changes in these sustained responses. Specifically, sustained response dynamics align with the Bayesian concepts of precision, an indicative of the confidence in predicted sensory input, pointing to the brain's adaptive mechanism to recalibrate its internal model based on the predictability of incoming stimuli (Zhao et al., 2024).

Zhao et al. (2024) provide direct evidence that sustained response may be the potential neural correlate of precision in auditory processing, a component not yet dissociated from neural recordings. As discussed in Section 2, gaining a nuanced understanding of precision's regulatory mechanisms could illuminate the pathophysiological underpinnings of various mental disorders and highlight potential therapeutic interventions. Consequently, one objective of this thesis is to dig deeper into the nature of these neural dynamics and provide insights for the underlying mechanism (see Chapter 3 and Chapter 5).

1.3.4. The Neural Substrates of Auditory Regularity Encoding

In addition to understand the statistical representations manifested by the sustained neural responses, Barascud and colleagues also probed the neural substrates that underpin these dynamics, particularly concerning the brain's detection of emerging patterns, where the neural responses evoked by regular patterns begin to diverge from those triggered by random sequences (see **Figure 1.1**), fMRI and MEG source localisation techniques pinpoint a critical network (Barascud et al., 2016). This network encompasses the auditory cortex, frontal cortex, and notably, the hippocampus.

The auditory cortex is apparently engaged during the auditory tasks, and its role of being a novelty detector has been documented in a wealth of literatures (Heilbron and Chait, 2018). However, the precise role of frontal cortex contributes to auditory processing and regularity encoding remain inadequately understood. Given the well-documented roles of the frontal cortex in cognitive control, attentional regulation and working memory, it is plausible that these cognitive functions are intricately involved in the auditory processing (Miller and Cohen, 2001). Empirical evidence indeed indicate its sensitivity to auditory deviants and auditory contextual change (May and Tiitinen, 2010; Näätänen et al., 2012; Paavilainen, 2013).

For instance, work by Doeller et al. (2003) investigated pre-attentive auditory deviance detection in prefrontal cortex, using oddball paradigm. In their experiment, standard tones at 500 Hz were contrasted with deviants at 667 Hz (small), 833 Hz (medium), and 1000 Hz (large). The findings showed distinct activity patterns in the right prefrontal cortex, especially noticeable during the processing of smaller pitch deviants. Particularly, the fMRI data demonstrated that the right prefrontal cortex showed increased activation when the deviants were less identifiable, an indicative of a specific role in contrast enhancement. Such pattern suggests that the prefrontal cortex enhances the sensitivity of the auditory detection system. This is particularly important under challenging conditions where auditory stimuli are subtle and less distinct. The observed activations show that the prefrontal cortex plays a significant role in top-down modulation, which helps to prioritize and intensify the neural processing of subtle deviations, thereby enhancing the perceptual clarity of these auditory differences. (Doeller et al., 2003)

In addition to the change detection, emerging evidence suggest that the prefrontal cortex plays a crucial role in detecting global deviations, highlighting its importance in interpreting, and responding to the auditory contextual inputs. Specifically, the study by Uhrig et al. (2014) explored the role of the prefrontal cortex in detecting global auditory deviants in the monkey brain. The group utilised the local-global auditory paradigm to test how monkey brain responds to hierarchical auditory deviations. The fMRI results indicated that while local deviants elicited a mismatch response predominantly in the auditory cortex,

global deviants activated a broad frontoparietal network, including significant involvement of the prefrontal cortex. The study found that the prefrontal cortex, particularly prefrontal areas 8A and the dorsal mid-cingulate, was critically involved in processing global deviations that require higher-order functions like integrating and comparing sequential information across time. (Uhrig et al., 2014)

Dürschmid et al. (2016) further addressed similar questions and provided empirical evidence in a study involving humans. The researchers recorded Electrocorticography (ECoG) signals from epilepsy patients to test the role of the pre-frontal cortex in detecting global deviants. In addition, they also examined whether the brain response differentiates between predictable and unpredictable deviations. The experiment used sequences of five tones for each trial, with deviants occurring either predictably every fifth tone or unpredictably within the sequence. Their results discovered that besides significant activation of the frontal cortex during the processing of global deviants, the heightened high gamma activity was observed in the frontal cortex in response to unpredictable auditory deviants compared to predictable ones. This finding demonstrated the crucial role of gamma band in the frontal cortex in signalling high-level auditory prediction errors. In contrast, the temporal cortex showed less selective responses, indicating a more generalized response to auditory deviations, regardless of their predictability. The study suggested that the selective response of the frontal cortex might be involved in functions such as updating internal model and recalibrating neural predictions in response to unexpected sensory input, as theorised by predictive coding (Dürschmid et al., 2016). Overall, these findings are consistent with Barascud et al. (2016), where the frontal cortex is involved in auditory regularity detection.

Apart from the frontal cortex, the hippocampus, traditionally recognized for its primary function in memory, also contribute to the neural network underlying auditory regularity detection (Barascud et al., 2016), though it is not typically observed in auditory tasks such as those with oddball paradigm, which generally emphasises sensory and attentional processes. The association between the hippocampus and auditory regularity detection could be due to the hippocampus's role in memory functions, necessary for integrating and retaining complex auditory sequences.

According to the IdyOM model, effective prediction of incoming sounds relies on the ability to memorize dependencies within sequences, aligning with the hippocampus's known functionalities. For example, early research by Kumaran and Maguire (2006) designed fMRI protocol to explore the hippocampus's role in detecting associative mismatches within visual object sequences. Participants were presented with sequences of objects where certain objects were expected based on prior exposures. These sequences were then altered in subsequent presentations to introduce associative mismatches — deviations from the expected sequences that were specifically designed to challenge the participant's memory-

based predictions. For example, if a sequence usually presents objects A, B, C, and D in that order, it might be changed to A, B, C, and X in a subsequent round. This creates a mismatch between the expected D and the actual X. Their fMRI results demonstrated a significant enhancement in hippocampal activation during conditions of associative mismatch. This suggests that the hippocampus is engaged in identifying when expected sequences are not followed, indicating its capability to learn dependencies within a sequence and actively compare incoming data against stored memories to signal 'prediction error' like mismatch. (Kumaran and Maguire, 2006)

Dimakopoulos et al. (2022) expanded the investigations on auditory working memory for processing auditory sequences. Specifically, they analysed how the hippocampus interacts with other brain regions to process and recall auditory sequences in verbal working memory. Using hippocampal local field potentials and electrocorticography (ECoG) recordings, the study examined the functional dynamics between the hippocampus and auditory cortex during verbal memory tasks. Participants were engaged in a modified Sternberg working memory task involving sets of consonants. The stimuli used consisted of a set of eight consonants where the central four, six, or eight letters were the specific memory items for each trial, distinguished by their set size. The task required memorizing and later recalling these sets, which assesses the participants' ability to handle varying memory loads. The study found that during the encoding phase, information primarily flowed from the auditory cortex to the hippocampus, particularly in the theta frequency range (4-8 Hz). Notably, during the maintenance phase—when participants were actively trying to retain and manipulate memorised sequences—the information flow reversed. This shift was indicated by the hippocampus predicting activity in the auditory cortex, suggesting its engagement in replaying and organising memory content for later recall. The authors noted that a higher memory load significantly altered neural firing patterns and functional connectivity, especially during the maintenance phase of the working memory task. This change was evident in the increased directional information flow from the hippocampus to the auditory cortex. The study demonstrated that the hippocampus not only plays a role in storing auditory working memory but also actively participates in processing and replaying sequential information within the working memory processes. (Dimakopoulos et al., 2022)

Further to this, the intracranial study by Borderie et al. (2024) provides a detailed analysis of the neural mechanism behind the maintenance of the auditory sequence in short-term memory. The study focuses on the role of the hippocampus, examining cross-frequency coupling in cortico-hippocampal networks. Piano tones were used in a short-term memory (STM) task for epilepsy patients, who were presented with sequences of 250-ms-long tones varying in memory load (3 or 6 tones) and silent retention periods (2 s, 4 s, and 8 s). Auditory STM was assessed by comparing two sequences, separated by a silent period, which were either identical or differed by one tone. The results demonstrated that stronger

theta-gamma phase-amplitude coupling in network of superior temporal sulcus, inferior frontal gyrus, inferior temporal gyrus and hippocampus, was associated with better performance on memory tasks, which suggests that such coupling plays a functional role in short-term memory retention. Notably, the activity observed in the network significantly predict memory performance at the individual trial level within participants, an indicative of the predictive capacity of this neural process. (Borderie et al., 2024)

In summary, insights from the above studies suggest that the hippocampus is not just a passive storage machine but is actively engaged in dynamic sensory information processing through complex neural interactions. Such active involvement sheds light on the underlying hippocampal mechanisms critical for auditory regularity detection, as initially proposed by Barascud et al. (2016). It raises the hypothesis that the hippocampus may function as a high-level predictive machine and contribute to sustained neural responses during auditory pattern detection tasks. Building on the foundation laid by Barascud's study, this PhD thesis introduced silent intervals between sounds to extend the duration of auditory patterns. Such modification aims to present a greater challenge to auditory memory systems, thereby enhancing our exploration of how memory processes contribute to the identification of auditory regularities.

1.4. Aim of This Project

The literature reviewed thus far have provided compelling evidence that the auditory system's sensitivity to patterns is intrinsically linked to memory functions. To decipher the nature of the neural correlates of regularity encoding, as demonstrated by sustained responses (Barascud et al., 2016; Southwell et al., 2017; Zhao et al., 2024), understanding the cognitive and perceptual interplay within this process appears to be the important aspect. This PhD thesis investigated several questions based on this topic: What types of information does auditory memory utilize to behaviourally analyse patterns within sound sequences? Does this memory integration in pattern extraction depend solely on the duration of sound sequences, or it adaptively monitor the item-wise information? How does those sensory information are represented in the brain? Is the memory mechanism supporting this process perceptual specific, or it is interacted with the cognitive functions? This thesis aims to answer these questions by utilising behavioural experimentation, along with MEG and EEG techniques.

2. Chapter 2: Unravelling the Interplay of Duration and Information boundaries in Rapidly Unfolding Sound Pattern Detection: Insights from Behavioural Examination

2.1. Introduction

In natural auditory scape, the arrival rate of sound streams can vary, even for the identical sources. This variability is exemplified by species-specific communication signals, such as bird songs (Podos et al., 2004), insect calls (Baker et al., 2019), and mammalian vocalisations (Jürgens, 2009), which all display unique temporal patterns. These differences underscore the necessity for organisms to adapt to the varying speeds of auditory cues to effectively process crucial environmental information, such as the presence of predators, prey, and food sources. The brain, central to auditory perception, is believed to continuously analyse statistical patterns within dynamic sensory signals across various auditory dimensions (Skerritt-Davis and Elhilali, 2021) and temporal scales (Fitzgerald and Todd, 2018). This ongoing analysis is critical for auditory sensory processing, as creating coherent concepts relies on the sequential unfolding of auditory events. For instance, in the realm of speech comprehension, integrating linguistic elements across different temporal scales—words, sentences, and paragraphs—is essential for forming a thorough understanding of the conveyed message (Diehl et al., 2004).

A fundamental question arises from these phenomena: how do listeners process and understand auditory stimuli that unfold over various time scales, and how does the brain integrate these signals over time to form a coherent perceptual experience? Furthermore, we must explore what specific types of information the brain is encoding during this process. The answer likely hinges on the mechanisms of memory, as local memory is believed to be interconnected with information processing (Hasson et al., 2015). Due to its limited capacity, the brain cannot encode every detail of an auditory signal. Instead, it is more likely that the brain employs a selective encoding strategy, prioritizing information that is essential for the organism's adaptation to its environment. This raises additional questions: what specific information within the auditory stream is deemed necessary to be tracked and retained, and for how long the brain consider retaining temporarily?

First, it is impossible for the memory to retain the information for infinite time. Empirical findings in experimental psychology support the view that auditory short-term memory possess a restricted temporal capacity (Näätänen et al., 1989; Winkler & Cowan, 2005). Such limitations are thought to stem from cellular level biophysical constraints, including the weakening of synaptic connections over time (Hardt et al., 2013). Classical memory studies employing two-stimulus comparison tasks (Cowan, 1984) have consistently demonstrated that participants are tasked with retaining an initial sound for a duration typically lasting only a few seconds. Any extension of this retention interval invariably results in a decline in comparison performance (Cowan et al., 1997).

Relevantly, McDermott and colleagues examined auditory representations in response to sound texture excerpts of varying lengths. They proposed that textures are rich in details, and it is more ecological for the brain to encode the sound by time-averaged statistics. To test this hypothesis, the team modified texture details and statistics in both short and long excerpts and evaluated listeners' discriminative abilities. As a result, it was found that listeners could distinguish short sound texture details effectively but struggled with time-averaged statistics. However, for longer excerpts, their ability to discern texture details decreased, and their proficiency in identifying differences in time-averaged statistics significantly improved (McDermott et al., 2013). Though memory is indeed temporarily limited, McDermott et al. (2013) provided evidence that temporal constraint does not preclude the brain's ability for encoding spectral temporal details of the auditory signal when it is relatively short. In other words, the brain appears to be capable of integrating and retaining distinctive features adaptively and comprehend sensory signals that vary in temporal details within certain range of temporal frame.

In fact, previous studies using tools like fMRI (Hasson et al., 2008; Lerner et al., 2011; Stephens et al., 2013) and EEG/ECoG (Lü et al., 1992; Honey et al., 2012), have provided insights that neural activity operates at various temporal scales, matching the stimulus inputs' presentation rate. For example, one recent research investigated how humans predict future events based on their past sensory experiences, particularly in situations where there are variations in the speed of sensory input. Participants in the study listened to sequences of pure tones presented at different rates: fast (150 ms per tone), medium (300 ms), or slow (600 ms). The recorded MEG data revealed that the brain's ability to anticipate upcoming information depends on integrating consistent amounts of tonal information, regardless of the rate at which it is presented (Baumgarten et al., 2021). Similarly, in a study on human speech processing, researchers investigated how the brain adapts to changes in speech rate. Using fMRI and intracranial EEG, the study examined neural responses to an auditory narrative presented at various rates. It was found that neural responses in early sensory processing auditory regions, as well as linguistic and extra-linguistic brain areas, could be temporally rescaled when speech was slowed down by 1.5

to 2 times or sped up by 0.75 times. However, this phenomenon started to break down for stimuli presented at double the speed, resulting in reduced intelligibility. This implies that flexible time scaling only occurs within specific time intervals. Once the speech rate exceeds a certain point, both duration and information can influence the perception (Lerner et al., 2014).

However, it is known that the processing of speech (Lerner et al., 2014) or slow stimuli (Baumgarten et al., 2021) likely involves the coordination of multiple neural circuits. This extends beyond early auditory sensory processing and includes higher cognitive functions like anticipation (Lee et al., 2021) and deliberate reasoning (Karlaftis et al., 2019), which complicate the results interpretation. To focus on the mechanism of the earlier stage, some studies have utilised rapidly unfolding tone sequences to investigate the automatic nature of auditory sensory processing.

For example, Watson and colleagues (Watson et al., 1990) firstly discovered that disruptions in fast tone patterns lasting 500 milliseconds and containing 10 items could be perceived. The study created random sequences of N tones by evenly distributing a fixed frequency range (300-3000 Hz) into N intervals on the logarithmic axis. Participants were asked to compare two patterns that differed by only one frequency element. Tests for discrimination ability were conducted at various levels of N and different tone/pattern durations. As a result, it was revealed that the total duration of the pattern and the duration of tones had minimal impact on the comparison task. Instead, the number of items (N) that made up the pattern accounted for the majority of the observed variance, suggesting the limited informational capacity in short-term memory.

Relevant evidence from behavioural study by Jaunmahomed and Chait (2012) also provide similar insights through how individuals retrospectively timestamp events. Participants were asked to decide whether a light flash occurred before or after the transition from a random tone sequence to the repeating tone pattern. Interestingly, the study found that the perceived timing of events was not at the point of detection, but slightly earlier – by about one cycle. Moreover, when the duration of tones was reduced from 100 to 50 milliseconds, this effect was halved. This suggests that the memory representation used to determine the start of a regular pattern is not tied to a specific duration, but to the number of available tones that are internally represented (Jaunmahomed and Chait, 2012).

Core question: Is the Memory that Supports the Pattern Discovery Dependent on Duration or Information

In addition to the behavioural perspective, neural correlates also offer objective evidence. Barascud et al. (2016) utilised MEG to further investigate how the human brain

automatically discovers the emergence of the sound pattern. Similar to Jaunmahomed & Chait, 2012, the stimuli consisted of rapid tone-pip sequences, with each tone lasting 50ms. The experiment included REG (regular) pattern with different levels of pattern complexity (Rcyc, the number of frequencies that make up the repeating pattern), specifically 5, 10, and 15 (REG5, REG10, and REG15). The duration of the patterns varied in terms of the size of the alphabet used in each condition, and the frequencies were always chosen from a pool of 20 frequencies. As a comparison, each REG signal was paired with a RAND (random) signal, which consisted of the same subset of frequencies (RAND5, RAND10, RAND15, RAND20, only RAND20 was plotted in this study) but presented in a random order.

Figure 2.1 illustrates the neural response from stimulus onset to offset for REG5/10/15 and RAND20 conditions. The experiment revealed that the amplitude of the sustained response is modulated by the predictability of the stimulus sequence. Specifically, REG5 > REG10 > REG15 > RAND20, indicating a larger amplitude with higher predictability. Interestingly, the time latency of the DC change, which is assumed to reflect the point of pattern discovery, increased with pattern complexity (Rcyc). The earliest DC change was observed in REG5, followed by the other conditions. The Ideal observer model (IdyOM) suggests a consistent requirement of three or four additional tones for pattern detection, regardless of informational complexity. However, brain responses from humans only mimic ideal observer performance for REG5 and REG10, with growing sluggishness for larger Rcyc sizes. This observation is also correlated with the DC amplitude effects: REG5 and REG10 show ideal observer performance and exhibit the same amplitude, while REG15 has a slightly lower amplitude. Since the sequences are too rapid to be consciously tracked, the study proposed that the auditory system needs to maintain and continuously update the sensory representation of the sound input to assist the pattern recognition. The decision regarding the emergence of the pattern is then made based on accumulating enough sensory evidence. The authors suggested that the decrease in performance after Rcyc of 10 might be attributed to the limited capacity in memory. Sound information that exceeds the memory capacity becomes non-retrievable, requiring access to the alternative memory strategy to support the belief update. This results in the delayed DC shift in more complex patterns (REG15). However, one confounding factor in this study is that when the pattern complexity increases, the duration of the pattern also increases; therefore it cannot be determined whether this constraints depends on duration or the number of tones as they both varied together in this study (Barascud et al., 2016).

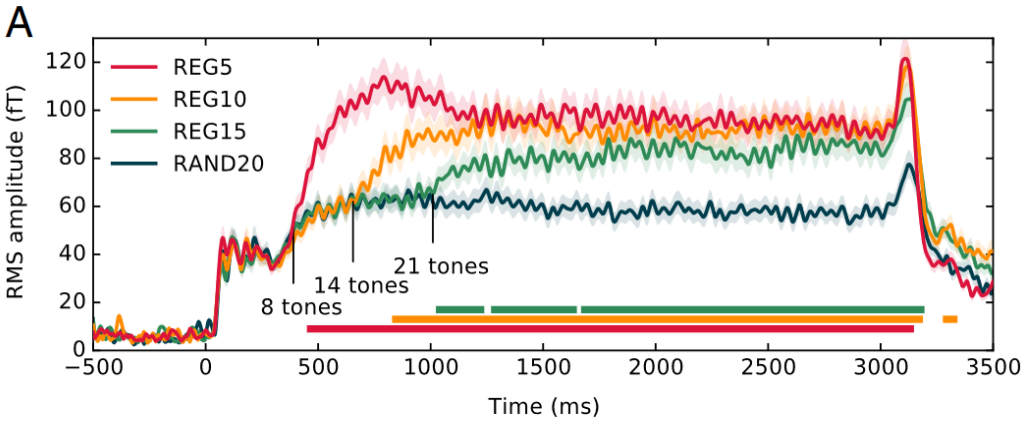

Figure 2.1. Group root mean square (RMS) of brain responses to REG5, REG10, and REG15 conditions, along with RAND20, are displayed. The entire stimulus epoch, from stimulus onset (t = 0) to offset (t = 3,500 ms), is plotted. Intervals where a repeated measures bootstrap procedure indicated significant differences between each REG condition and RAND20 are marked with a line beneath the brain responses. (Barascud et al., 2016)

Harrison and colleagues (Harrison et al., 2020) proposed that the pattern duration of auditory sequences in Barascud et al.'s study might be within the temporal boundary of echoic memory (Nees, 2016; Winkler & Cowan, 2005). It was argued that the limited capacity for storing number of information could be a factor. To test the hypothesis, the team conducted a behavioural experiment, in which the participants in the experiment were asked to identify when a random tone sequence transitioned into a regular pattern. The response time for pattern recognition was compared under two conditions. Both conditions maintained a pattern duration of 500 ms but differed in the number of tones per cycle (Rcyc); one had 10 tones lasting 50 ms each, while the other included 20 tones each lasting 25 ms. The results revealed that an increase in the number of tones substantially increased the tonal information participants needed to detect regularity. This was true even when the duration of each cycle was kept constant and fell within the hypothesised echoic memory boundary. Those observations suggest that the memory buffer which supports listeners to detect the regularity are indeed influenced by informational constraints.

This finding was also supported by McDermott et al. (2013), who created 'cocktail party' textures by overlaying recordings of groups of speakers. Listeners were tasked with determining which of three sound excerpts, all sampled from the same signal, was distinct from the others. The signals contained varying numbers of different speakers. The results demonstrated that short duration excerpts (50ms) were easily distinguishable irrespective of number of speakers. However, performance became significantly worse for longer

durations (2500ms) depending on the conditions, showing an interaction between duration and the number of sources. Their subsequent experiment involved generating random sequences of drum hits ranging from sparse to dense, further hinting at an interaction between duration and hit density. Those results indicate that once the information flow exceeds a certain temporal threshold of the memory, penalty of informational limits arise (McDermott et al 2013).

2.1.1. The Motivation Behind the Study

Harrison et al. (2020) analysed a small dataset and included a limited range of stimulus conditions in their study. Meanwhile, it is important to note that adjusting the tone length to manipulate the cycle duration can potentially enhance the encoding of sensory signal and reduce the effects of decay in sensory trace. Therefore, it would be unfair to solely compare the results based on the controlled cycle duration length (Harrison et al., 2020).

To enhance our understanding about this topic, this study expands upon prior research paradigms by tasking participants to distinguish the transition from a random sequence of tones (RAND) to a structured, repeating pattern (REG). In alignment with established research, pure tone patterns as stimuli were employed. This choice is advantageous because tone-pips encapsulate essential auditory elements and offer precise control. Additionally, the rarity of pure tone patterns in natural settings reduces the likelihood that the observations are influenced by participants' long-term memory, ensuring a focused investigation of the specific auditory processes. Via manipulate the informational and temporal parameters of the pattern, this study investigated how the ability to detect the transition changes as the function of pattern complexity (Rcyc, the number of tones/frequencies that make up the pattern) or the duration of the pattern. Response time (RT) was utilised to evaluate task performance, since RT is a commonly measured metric in the perceptual decision-making process. The variability in its distribution can offer valuable insights into how sensory evidence is gathered over time and how information is internally represented (Glickman and Usher, 2019). RT in this study is defined as the duration listeners take to perceive the pattern emergence, once adjusted for the baseline response latency necessary to detect a straightforward tone change (STEP). Critically, the response time to transitions in the STEP stimuli is used to estimate the response time to a simple, computationally low-demanding pitch change. This considers any effects of hardware latency, the time it takes for the change to reach awareness, the time to trigger and complete the motor response, and the subject's overall attentiveness level. Therefore, RTs measured in pattern (RANDREG) detection in this study were all corrected by subtracting the STEP RT.

To demonstrate the significance of RT and its correlation with memory capacities, this study utilised the PPM decay model (Harrison et al., 2020). This model aids in understanding the consequences of memory constraints on the temporal dynamics of sound sequence processing. **Figure 2.2**A displays the decay parameters used in the model. Information (transition probabilities) stored in the echoic memory buffer can be recalled with high precision. However, memories that leave the buffer over time transition into short-term memory, beginning with a 0.1 weighted memory strength and decays exponentially (see the demo displayed in **Figure 2.3** explain how the pattern can be theoretically detected considering the decay process). Same principle was applied to long-term memory decay after the short-term memory phase. **Figure 2.2**B demonstrates how the model process REG10 or REG20 with three buffer sizes. When the buffer is reduced to 500ms, the detection of REG10 appears to lag compared to the other two buffer sizes (see **Figure 2.2**B), and this is also true for REG20 detection.

Figure 2.4 illustrates how the model processes RANDREG10 (pattern duration = 500ms) and RANDREG20 (pattern duration = 1000ms) with corresponding decay parameters. The y-axis shows the information content, which quantifies the level of surprise after the tone at that position is observed by the model. Less surprise indicates that memory has been formed, and the model can make more confident prediction. In this example, I modelled the memory buffer with temporal constraints of 0.5, 0.95, and 1.5 seconds. The model suggests that when the pattern completely fits within the memory buffer's temporal frame (i.e., 1.5 seconds), the information content drops to the hypothesised detection threshold at the same time point for both REG10 and REG20. However, when the memory buffer cannot cover the length of the pattern (i.e., 950ms for REG20), the model needs more processing time to reach the detection threshold for REG20 compared to REG10.

In essence, the model illustrates how response times are likely to vary based on assumed memory's temporal capacities in pattern detection tasks. When the model include temporal constraints alone, larger temporal buffer sizes allow for efficient detection, resulting in same response times regardless of pattern duration (e.g., bottom plot in **Figure 2.4**). Conversely, smaller buffer sizes lead to slower response times that are dependent on pattern duration. This occurs because the high-fidelity echoic memory buffer cannot fully store pattern information, necessitating the gathering of more sensory evidence from other form of storages over time.

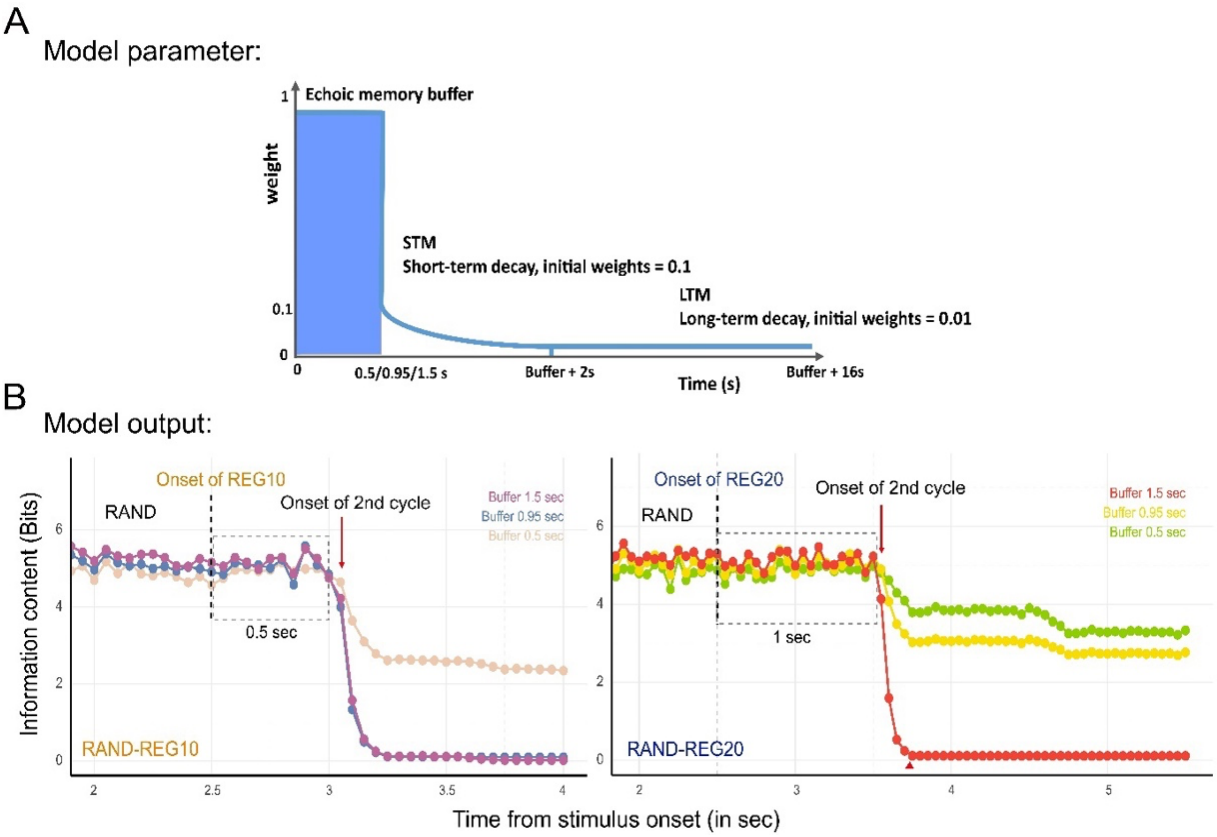

Figure 2.2. In the context of PPM decay model (see methods section), this study examined how the echoic memory buffer size can affect the information content dynamics over time. (A) The curve in the PPM decay model represents memory decay. The buffer, shown as a blue square, symbolises high-fidelity echoic memory with a duration of 0.5, 0.95, or 1.5 sec in the model. This is followed by a short-term memory phase that decays exponentially, starting from 0.1 of the buffer's weight and lasts 2 sec. The long-term memory phase follows afterward. (B) The graph displays a model simulation of RANDREG10 (pattern duration of REG10 = 0.5 sec) and RANDREG20 (pattern duration of REG20 = 1 sec) detection. The REG pattern starts after 2.5 sec. The model estimates the information content (surprise) of each tone based on the experience of previous tones (represented on the y-axis). The memory decay parameters influence the estimation of information content by assigning specific weights to past experiences. The model indicates that the information content starts to decrease after observing one REG cycle in both pattern conditions. When the echoic memory buffer (1.5 sec) fully encompasses the pattern, the model only observed about 4 tones to reach the baseline (strong model). Yet, for memory buffers (0.5 sec in REG10, 0.5 and 0.95 sec in REG20)

that cannot cover the entire pattern duration or just match it, the model presents a slower decline in slope, and needs more tonal information (more evidence, interpreted as extended detection time) to reach a state of reduced surprise. (Harrison et al., 2020)

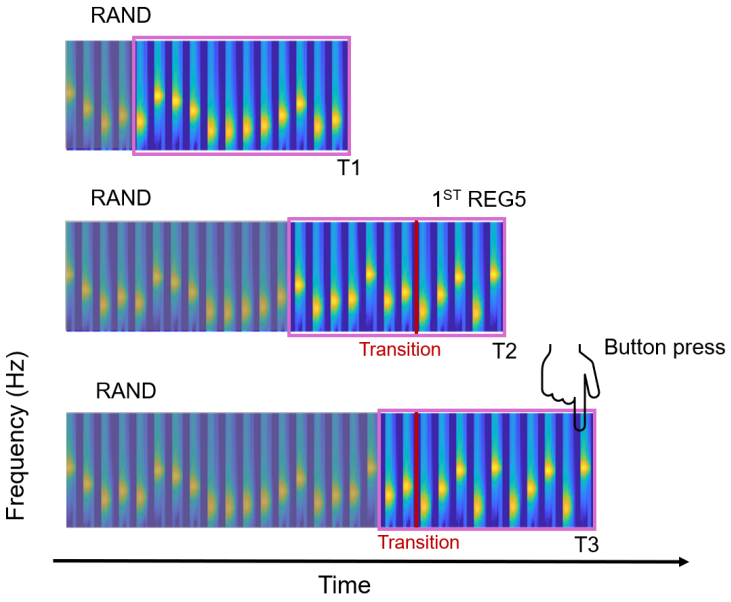

Figure 2.3. **This demo illustrates how a pattern becomes theoretically detectable throughout the unfolding of a sound sequence over time (this example plots the spectrogram of RANDREG5), assuming the limited temporal capacity of the memory buffer as proposed by the PPM decay model.** The pink frame signifies the memory buffer, while the grey shaded area represents the previously heard information that has fallen out of the memory buffer over time and undergone decay. When listeners are processing the sound, at the T1 time point (top), the transition from RAND to REG has not yet occurred. At the T2 point (middle), the red line indicates the transition from RAND to REG5, but since the REG5 has not started repeating, it is not theoretically detectable. By the T3 time point (bottom), the REG5 starts to repeat and becomes detectable within the second cycle. Listeners were instructed to press the button when they perceive the transition.

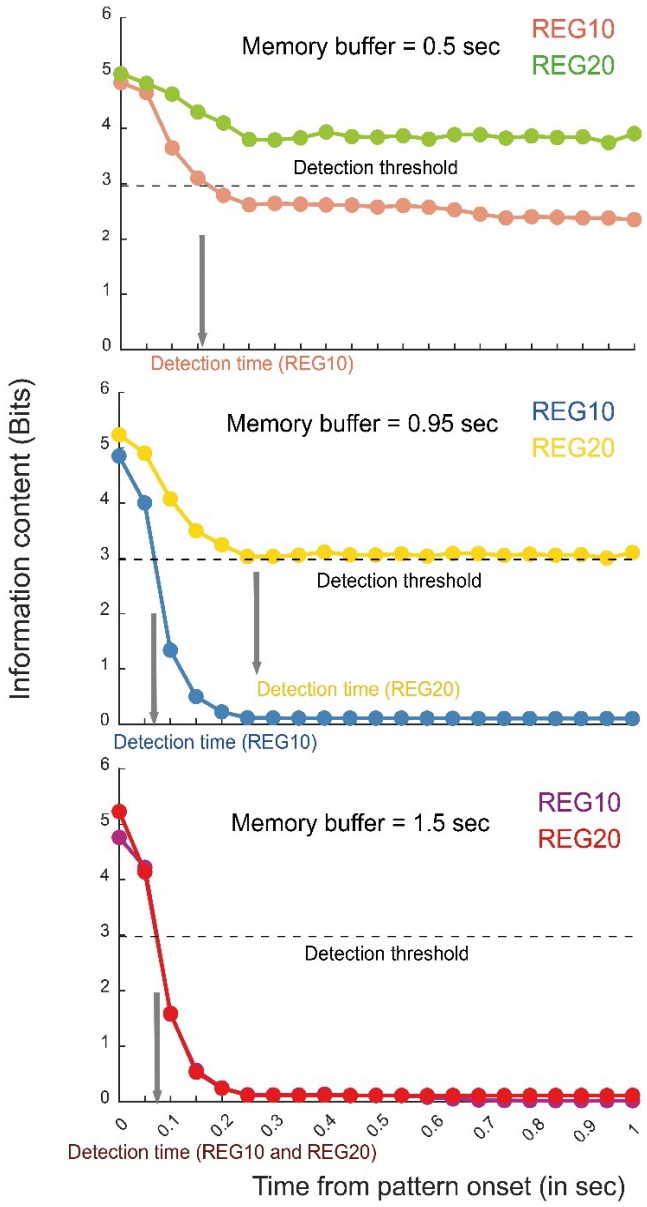

Figure 2.4. Comparison of the model output between REG10 and REG20 in different temporal capacities of the echoic memory buffer. The plot shows the output that starts from the onset of the first REG cycle, the black dashed line indicates the arbitrarily defined detection threshold. The duration of REG10 is 0.5 sec, and the duration of REG20 is 1 sec. The top figure shows that when the memory buffer (0.5 sec) equals to the duration of REG10 and is shorter than REG20, the model's surprise exhibits a gentler decrease slope in REG20 compared to REG10. In this case, REG10 takes more time to reach the detection threshold. However, due to memory decay, REG20 is not detectable within cycle 2. The middle figure

demonstrates that the detection threshold is reached just as REG10 begins to repeat, with the memory buffer at 0.95 seconds. However, REG20, due to its longer duration and increased susceptibility to decay, takes more time to reach the same detection threshold. The bottom figure illustrates a 1.5 sec memory buffer, which is long enough for both pattern conditions to reach the detection threshold at the same time after the patterns start to repeat.

2.1.2. Aim & Hypotheses

The PPM-decay model, while insightful, has limitations due to its dependence on specific assumptions about memory capacity, types of constraints (i.e. temporal constraint or information constraint), and the processes of encoding and decay. These assumptions might not fully encapsulate the complexity, variability, and uncertainty that characterize human echoic memory processes, particularly concerning the diversity of information encoded and the elusive understanding about the mechanisms of encoding. However, these limitations inspired the current study, which seeks to deepen our understanding of the auditory memory mechanisms involved in pattern detection. By investigating how human response times of auditory pattern detection are influenced by constraints related to information or duration, the research aims to enhance the understanding of dynamics of echoic memory. Two hypotheses are proposed in this study.

The hypothesis 1 suggests that auditory memory integrates sensory signals within a fixed timeframe. Instead of sequentially encoding unfolding items, the brain integrates chuck of information in a set temporal window. Consequently, irrespective of the number of discrete items that are included in the pattern, response time should remain statistically indistinguishable as long as the pattern duration remains the same (**Figure 2.5**B).

Alternatively, human listeners might adaptively integrate discrete informational units. The hypothesis 2 is tested against the IdyOM model (**Figure 2.5**A) which solely considers item-wise information, free from constraints (Pearce, 2005). This hypothesis suggests that human participants monitor the transition probabilities of individual tones. In terms of IdyOM, it requires fixed amount of information to detect the pattern, regardless of pattern complexity (Rcyc), provided the elements that consists of the pattern were selected from the same pool. From this assumption, this study proposes that the number of information/tones needed to detect the transition remains constant, regardless of pattern complexity. The detection threshold of the model in each condition was utilised to represent the boundary of performance in hypothesis 2 (**Figure 2.5**B).

It's important to note that all forms of memory inherently undergo temporal decay. The PPM decay model also reflects a longer detection time when this decay occurs (**Figure 2.2**B). To test the hypotheses without influence of this factor, in experiment 1, a pattern duration of 500ms was maintained, short enough to fall within the limits of echoic memory. The study varied pattern complexity (Rcyc = REG5, REG10, REG15, REG20), keeping tone length fixed at 25 ms, and manipulated Rcyc by introducing silent gaps between tones to maintain fixed pattern duration. If hypothesis 1 predicts the human performance, a relationship of $RT_{RANDREG5} = RT_{RANDREG10} = RT_{RANDREG15} = RT_{RANDREG20}$ is anticipated, to linearly transform the RT measured as milliseconds into number of tones (see Methods for how the linear transformation is performed), the hypothesised relationship will be $RT_{RANDREG5(number\ of\ tones)} = 2RT_{RANDREG10(number\ of\ tones)} = 3RT_{RANDREG15(number\ of\ tones)} = 4RT_{RANDREG20(number\ of\ tones)}$ (**Figure 2.5**B). Conversely, if the human listeners were adaptively integrating discrete items, an IdyOM model like relationship of $RT_{RANDREG5(number\ of\ tones)} = RT_{RANDREG10(number\ of\ tones)} = RT_{RANDREG15(number\ of\ tones)} = RT_{RANDREG20(number\ of\ tones)}$ is expected, and this is equivalent to a relationship of $RT_{RANDREG5} = 2RT_{RANDREG10} = 3RT_{RANDREG15} = 4RT_{RANDREG20}$, when $RT_{number\ of\ tones}$ are converted into $RT_{in\ ms}$.

In Experiment 2, this study aimed to further explore the memory process described in Experiment 1. Four experimental conditions orthogonalizing two stimulus dimensions were created: Rcyc (REG10 or REG20) and pattern duration (500ms or 1500ms). The objectives were twofold: to replicate results observed in Experiment 1. To investigate how an increase in pattern duration influence the RT variations of pattern detection, and which hypothesis can predict the human performance in slow sequence processing. Similarly, if hypothesis 1 predict the human performance, the study anticipate the relationship of $RT_{RANDREG10} = RT_{RANDREG20}$ or equivalent relationship of $2RT_{RANDREG10(number\ of\ tones)} = RT_{RANDREG20(number\ of\ tones)}$ in both duration conditions. Alternatively, hypothesis 2 will predict a relationship of $RT_{RANDREG10(number\ of\ tones)} = RT_{RANDREG20(number\ of\ tones)}$ in two duration conditions.

Furthermore, both Experiment 1 and Experiment 2 used a paradigm manipulating the duration of a silent gap (25 ms) between fixed tone-pips. However, it's essential to acknowledge that this silent gap may introduce factors affecting auditory perception. To address this concern, experiment 3 was designed with the same paradigm as Experiment 2, but with stimulus pattern duration altered by varying tone length.

Exp1, Hypothesis

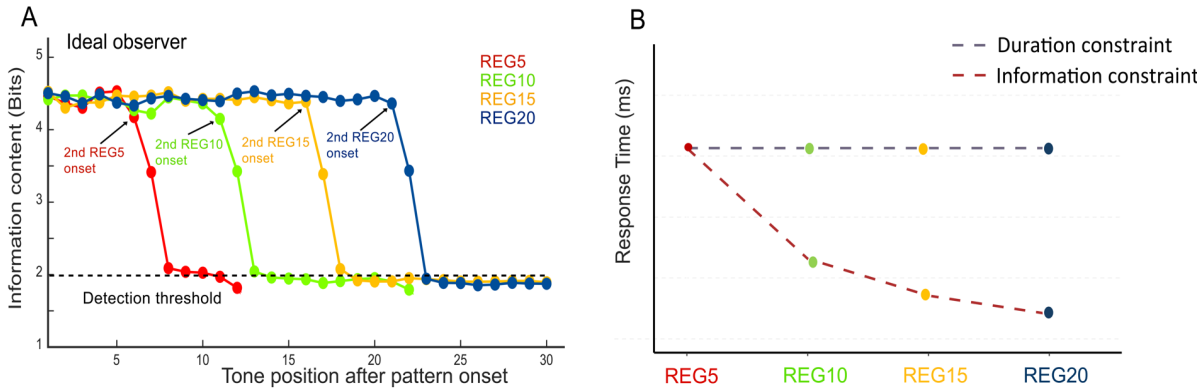

Figure 2.5. Experiment 1 hypotheses. (A)The ideal observer model served as the boundary for assessing the informational constraint of REG detection. Regardless of pattern complexity (Rcyc), the model can detect the pattern with just observing 3 or 4 tones within the second repeated pattern. The dashed line indicates the detection threshold, signifying successful prediction of observed tones. (B) Hypothesised response times (in ms) are delineated in relation to each constraint. Hypothesis 1: duration constraints are expected to yield uniform response times across conditions. Hypothesis 2: information constraints anticipate a consistent amount of information(tones). This information, when transformed into time linearly, exhibits a declining trend across conditions.

2.2. Experiment 1. Is Pattern Detection Ability Limited by Pattern Complexity

2.2.1. Methods

In this experiment, the primary goal was to investigate the impact of pattern complexity, denoted as Rcyc, on memory integration. To achieve a controlled environment for the study, the study standardised the presentation time for all patterns to a fixed half-second (500 ms) interval, complemented by a uniform tone duration of 25 ms for each sound. This uniformity in tone length was critical to ensure that the sensory system's encoding of frequency information remained consistent across all test conditions. In terms of the memory studies (Cowan, 2008), it is expected that this duration would be sufficiently within the limits

of echoic memory to prevent temporal decay from impacting the results, thus allowing to focus on the influence of informational complexity in memory integration process.

The experiment incorporated four distinct pattern conditions with varying Rcyc values, which represent the alphabet size of the pattern: REG5, REG10, REG15 and REG20 (see stimuli example in Figure 2.6). By adjusting the silent gaps between tone pips and the number of tones in each cycle, the study aimed to isolate the effects of pattern complexity (as determined by the number of tones the pattern comprises) on echoic memory processing. Two hypotheses are proposed: The first suggests a constancy in $RT_{in\ ms}$ across varying degrees of pattern complexity, contingent upon the assumption that the memory integrate information in terms of a fixed temporal frame. The second hypothesis is informed by PPM model (Pearce, 2005), which suggests a fixed number of information ensuring a confident pattern detection, the linear transformation of this representation into time will exhibit a decrease in $RT_{in\ ms}$ corresponding with an increase in Rcyc (Figure 2.5B). To transform $RT_{in\ ms}$ measured in milliseconds to the number of tones, the following linear transformation is applied in this study:

$$\mathrm{RT(number\ of\ tones)} = \frac{\mathrm{RT} \times \mathrm{Rcyc}}{\mathrm{Pattern\ duration}}$$

2.2.1.1. Participants

70 participants were recruited through Prolific (www.prolific.co) and completed this experiment. Of these, data from 11 participants were rejected due to reports of a noisy environment (see " data rejection criteria " below). Data from 7 participants were rejected due to failure to respond to STEP trials or because responses to STEP trials were too slow, and data from 2 participants were rejected due to extremely low d prime (d'). In total, 50 participants (19 females; average age 24 ± 4.44 years) were included in the following analysis. In addition, 33 participants did not proceed to the main task due to not passing the pre-determined performance threshold in the practice task, meanwhile, about 28% of participants who initially accessed the experiment but did not pass the headphone screen and therefore did not proceed further (Milne et al., 2020).

2.2.1.2. Stimuli

Stimuli (See Figure 2.6) consisted of sequences of 25-ms tone pips, which were gated on and off with 3-ms raised cosine ramps. The frequencies of the tone pips were randomly drawn from a pool of 20 values that were equally spaced on a logarithmic scale between 222 and 2000 Hz (12% steps; loudness normalised based on iso226). A new sequence was generated for each trial. RANDREG sequences included a transition from a random (RAND) to a regularly repeating cycle (REG) of tone-pips. Each REG sequence

consists of three REG cycles. The tone-pips that make up the REG pattern for each condition were randomly selected from the full pool (20 frequencies). REG conditions of 5 (REG5), 10 (REG10), 15 (REG15), and 20 (REG20) tones were included. Each pattern condition was presented in a separate block with a pseudo-random (Latin Square function implemented on Gorilla experiment builder) order for each subject. The cycle duration was fixed at 500 ms, and inter-tone intervals were manipulated to fit that duration. Therefore, REG5 contained silent gaps of 75 ms between tones; REG10 contained silent gaps of 25 ms; REG15 contained silent gaps of 8.33 ms; and REG20 contained no silent gaps (0 ms). The RAND portion of the RANDREG sequences were generated by randomly sampling from the full pool with replacement. They contained the same inter-tone intervals as the REG portion. Transition onsets were randomised between 2-3 s post onset to ensure that the transition time was not predictable. RAND sequences consisted of tone-pips arranged in random order, with each frequency occurring equi-probably across the sequence duration. In each block, RAND was matched with RANDREG in inter-tone interval in each sequence. Two control stimuli were also included: sequences of contiguous (no silent gap) tone-pips of a fixed frequency (CONT) that lasted 4000 ms, and sequences (Figure 2.6) with a step change in frequency partway through the trial (STEP: the change always occurred after 2000 ms). These were used to measure individuals' baseline response time to simple acoustic changes and at the same time served as 'catch trials' to assess task engagement.

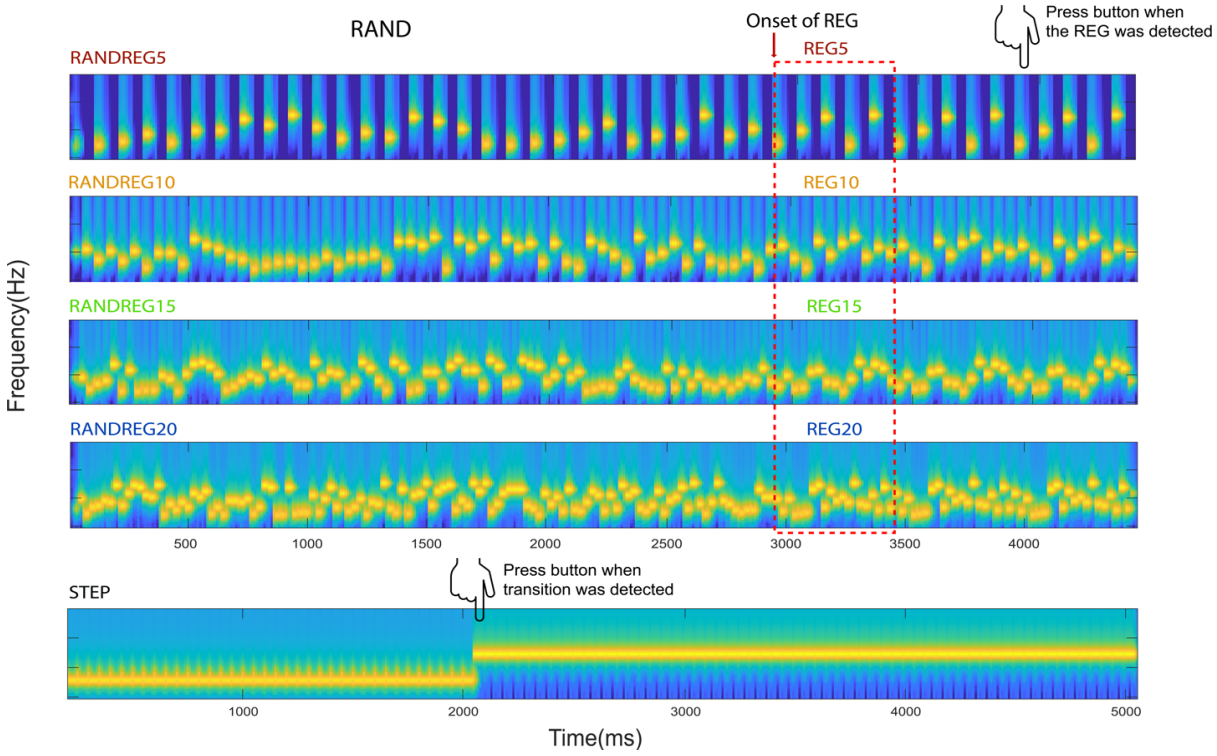

Figure 2.6. Spectrogram of example RANDREG stimuli of all Rcyc conditions in the short pattern duration (500ms). The spectrograms shown are from the RANDREG and STEP stimuli (target trials) used in experiment 1. These are auditory spectrograms, created with a filter bank of channels that are 1/ERB wide (Equivalent Rectangular Bandwidth; (Moore and Glasberg, 1983)). These channels are evenly distributed on an ERB-rate scale. To achieve a temporal resolution akin to the Equivalent Rectangular Duration (Plack and Moore, 1990), the channels were smoothed. Four stimuli feature a transition between a random and a regular sequence (REG5, REG10, REG15, REG20). The evenly distributed purple areas represent the silence interval placed between tone-pips, ensuring a consistent pattern duration of 500ms across all Rcyc conditions. Each tone lasts 25 ms. The onset of REG is indicated by left red dashed line and all target stimuli include three cycles of REG. Listeners were instructed to press the button as fast as they can once they perceived the pattern.

2.2.1.3. Procedure

The experiment was conducted online using the Gorilla Experiment Builder (www.gorilla.sc). Before the main experiment, participants completed a headphone screening task (Milne et al., 2020) to ensure they were using appropriate audio equipment. The main experiment was preceded by a volume adjustment stage. Participants heard a few sounds from the main task and were instructed to adjust the volume to a comfortable listening level. In the main experiment, participants were instructed to monitor for transitions (50% of trials) from random to regular patterns (RANDREG) and frequency changes in STEP stimuli and press a keyboard button as soon as possible upon change detection. The participants then received an explanation of the task and completed a practice session to become familiar with it. The main experiment was divided into four test blocks, each lasting 5-7 minutes, with one block for each REG condition. The order of the blocks was randomised for each participant. Participants were instructed to press a keyboard button as soon as possible once they detected a transition to RANDREG or a STEP change. Each test block consisted of 40 trials, delivered in random order. The block contained the following sequence types: 20 RANDREG, 20 RAND, 5 STEP, and 5 CONT. The main experiment lasted approximately 30 minutes. To encourage participants to concentrate on the task, feedback was provided on accuracy and speed at the end of each trial, similar to the previous work (Bianco et al., 2020). A red cross was displayed for incorrect responses, and a tick was displayed for correct responses. The colour of the tick was green if the responses were 'fast' (RT less than two REG cycles from REG onset or less than 500ms from the STEP change),

and orange otherwise. A small monetary bonus was given for each correct response, and the bonus was doubled for 'fast' responses. This served to encourage participants to respond as quickly as possible. (Bianco et al., 2021). The inter-block intervals were set to have a maximum duration of 2 minutes to keep the overall duration of the exposure equal across participants.

d' (d prime) serves as a general measure of sensitivity to patterns based on signal detection theory (Stanislaw et al, 1999). d' was calculated as: $d' = Z(\text{Hits}) - Z(\text{False Alarms})$. Hits were defined as responses occurring after the onset of the second REG pattern, while false alarms were responses to RAND trials or those occurring before the emergence of the REG pattern. A good d' suggests high sensitivity, and thus response times (RTs) are interpretable. The core analysis focused on RTs to the onset of regular patterns, where RT was defined as the time difference between the onset of the REG pattern or the STEP change and the participant's button press. The median STEP RTs computed per test block were used as a measure of the baseline latency of the response to a simple acoustic change and subtracted from the RANDREG RTs to yield a lower-bound estimate of the computation time required for change detection.

2.2.1.4. Data Rejection Criteria

Due to the online nature of the present experiments and associated reduced control over participants' environments, equipment, and engagement, it was important to implement a series of rejection criteria to make sure that data reflect true sequence tracking ability. Therefore, participants' data were excluded from all experiments following the below (A-priori determined) criteria:

(1) Failure on the Headphone screen: the task introduced by Milne et al. (2020) was used. Participants who did not pass the screening procedure did not proceed to the main experiment.

(2) Low performance in the practice run: To ensure that participants understood the task, a practice run of the pattern detection task (13 RANREG, 13 RAND, 2 CON, and 2 STEP) was delivered. Participants who scored below 60% in the practice task did not proceed to the main task. Our previous experience with similar stimuli in lab settings (Barascud et al., 2016; Bianco et al., 2020) suggests that the vast majority of young participants can achieve ceiling performance. This study, therefore, reasoned that those

online participants who performed below 60% are likely not sufficiently engaged with the task (i.e. distracted, not following instructions, etc).

(3) Of those participants who completed the full experiment, the data from those participants who failed to respond to STEP trials (allowing at most one miss per block) or whose RT to STEP trials fell above 2 STDEV relative to the group mean were rejected. Failure to respond quickly to the (easy) STEP trials indicated low task engagement.

(5) The exit questionnaire asked participants to rate the amount of background noise or interruptions they experienced during the experiment, with 0 indicating no noise and 10 indicating extreme noise. Data from participants who rated their environmental noise as more than 2 were excluded from the analysis.

(6) Importantly, to allow the study to quantify changes in performance as a function of Rcyc, it was critical that baseline performance was high. Therefore, data from participants whose mean d' (across conditions) was below 2 were not included in the analysis.

2.2.1.5. Statistical Analysis

Performance data were modelled by linear analyses of variance (ANOVA) implemented in ez package of R. When sphericity assumptions were violated, Greenhouse-Geisser adjustments was applied. Post-hoc t tests were used to compare performance differences between conditions across blocks and groups. Any value below the significance level of 0.05 is indicated as non-significant (n.s).

2.2.1.6. Modelling

The hypothesis-supporting illustrations, as depicted in Figure 2.2 and Figure 2.4, were generated using a memory-constrained version of the Prediction by Partial Matching (PPM) model. PPM is a form of Markov model adept at estimating the likelihoods of sequences of symbols by analysing the frequency of n-grams—sequences of 'n' items—within a set of training data. This method effectively smooths the transition between models of different orders by considering variable-length context histories.

The original PPM model (IdyOM), serving as a benchmark for the informational constraints hypothesis in this study, shown in Figure 2.5A, possesses an unbounded trial-

wise memory, retaining all items within trial with equal weight, regardless of their distance from the current event being modelled (Pearce, 2005). However, to reflect the human memory processes more accurately, Harrison et al. (2020) introduced a modified PPM with hypothesised memory phases and decay function. This 'PPM decay model' dynamically attenuates the influence of historical data over time using a customizable decay kernel. (Harrison et al., 2020) This decay kernel is triphasic: it commences with an echoic memory buffer that temporarily retains high-fidelity information, then transitions to a short-term memory (STM) phase where the weight decays exponentially from the buffer's weight to a baseline level over a certain duration and concludes with a long-term memory (LTM) phase where the weight diminishes exponentially, defined by its initial value and decay half-life.

For modelling, this study used tones each lasts 50 milliseconds (ms) to simulate the temporal precision required for the decay process. The tone sequences were generated using the same computational techniques as in the experiments (as detailed in the method/stimuli section of Experiment 1). Sequences of RANDREG10 (Rcyc = 10) and RANDREG20 (Rcyc = 20) were utilised for model inputs. A transition from random (RAND) to regular (REG) patterns was programmed to occur after every 50 tones. The sequence processing of the model was dynamic, with the likelihood of each tone being calculated in the context of previous sequences and considering the cumulative history of stimuli, thereby simulating long-term memory effects. To calculate information content, the model converted these probabilities using the negative logarithm to the base 2. The model's complexity was constrained to a maximum n-gram length of 5 symbols.

2.2.2. Results

In this experiment, participants were encouraged to detect STEP and RANDREG transitions as quickly as possible. The variations of response times (RTs) and its implications about the internal representation of sensory signals were of particular interest. Therefore, the study focused on analysing the RANDREG RTs, which were corrected by the STEP RTs. First, the response time of STEP detection (STEP RT) remained stable across blocks [$F(3, 147) = 1.359$, $\eta 2 = .007$, $p = .258$], suggesting that participants maintained a similar level of task engagement across the course of the experiment.

The average d Prime score (d'; see **Figure 2.7**A) was close to ceiling performance across all conditions which suggests the RTs are interpretable. A repeated measures ANOVA on d Prime data with Rcyc as a within-subject factor confirmed no difference between conditions [$F(3, 147) = 1.98$, $\eta 2 = .022$, $p = .12$]. This further confirmed that it is

reasonable to focus on interpreting response time (RT) to quantify how much information/time was required by human listeners to detect the repeating pattern. The response times (RTs) of effective transition, shown in **Figure 2.7**B and subsequent figures, represent the time taken to respond relative to the onset of the second cycle of the REG. The use of effective transition in this study aims to straightforwardly depict the processing time, as the pattern only theoretically becomes detectable in cycle 2.

The analysis of repeated measures ANOVA on RTs (in ms) revealed a significant effect of Rcyc [$F(3, 147) = 3.536$, $\eta 2 = .039$, $p = .016$], post-hoc t test indicated that RT in REG20 condition is shorter relative to REG5 [$t(1,49)=2.789, p = .003$] with mean differences about 41 ms; or REG10 [$t(1,49)=3.1318, p = .007$] with mean differences about 33 ms.

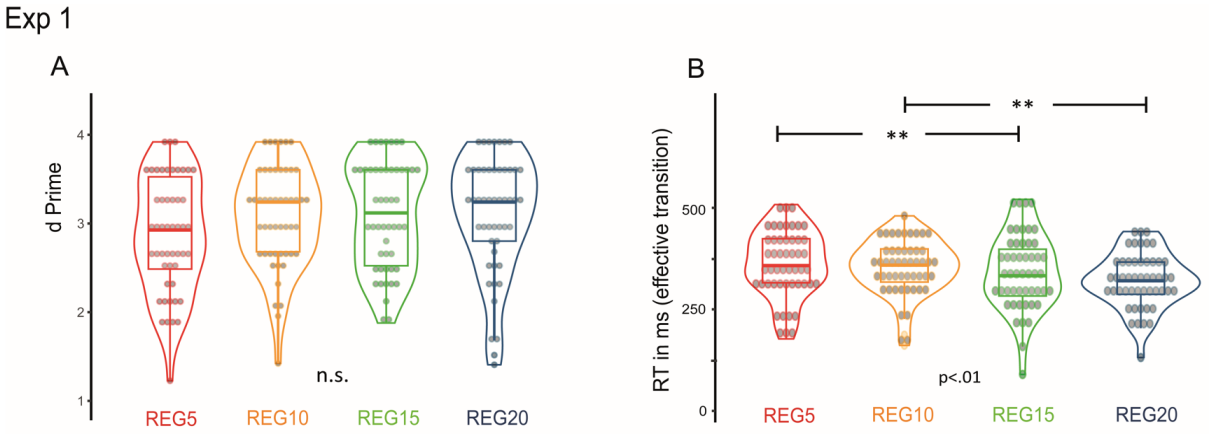

Figure 2.7. Behavioural performance. (A) d Prime distribution from experiment 1. Most participants exhibit high sensitivity to the pattern emergence, and no significant differences were observed across conditions. (B) Distribution of response times (effective transition) across conditions in experiment 1. Reduced RTs were observed when the Rcyc increased, with statistically significant differences between REG5 and REG15 or REG20 (paired t test, **<.01).

In the investigation of memory buffer and its role in pattern detection, two competing hypotheses were addressed as introduced above: one predicated on fixed temporal frames (Hypothesis 1: Duration constraints), and the other on a fixed quantity of information (Hypothesis 2: Information constraints) (see **Figure 2.5** B). Hypothesis 1 suggests that the auditory system integrates perceived signals within a constant temporal window. Therefore,

if Hypothesis 1 is correct, it would be expected that no significant variation in response times (RTs) for pattern detection tasks - predicting $RT_{RANDREG5} = RT_{RANDREG10} = RT_{RANDREG15} = RT_{RANDREG20}$ - since only the pattern complexity (Rcyc) changes while the pattern duration remains constant.

Conversely, Hypothesis 2 posits that pattern detection is incorporating a fixed number of informational units, irrespective of the complexity of the pattern. Should this hypothesis hold true, a constant number of tones would be necessary for listeners to identify the pattern across various conditions in terms of PPM model prediction. This would suggest RTs in terms of tones to be equivalent, thereby $RT_{RANDREG5(\text{number of tones})} = RT_{RANDREG10(\text{number of tones})} = RT_{RANDREG15(\text{number of tones})} = RT_{RANDREG20(\text{number of tones})}$.

Interestingly, the results from experiment 1 point to a nuanced interaction between duration and pattern complexity. If duration were the sole factor, $RT_{in\ ms}$ would consistently align across varying conditions. However, an observed reduction in $RT_{in\ ms}$ with an increase in Rcyc suggests that pattern complexity also plays a role. As illustrated in **Figure 2.8**, the group $RT_{in\ ms}$ (effective transition – corrected by the duration of first cycle of REG) vary significantly across conditions. The grey dashed line in these figures represents the trajectory anticipated by the duration constraint hypothesis, which would predict unvarying $RT_{in\ ms}$ if temporal duration were the only determinant. The red dashed line signifies the hypothesis centred around informational constraints, operating under the assumption that the brain functions akin to an IdyOM (**Figure 2.5**A), which processes tonal information on an item-wise basis. Notably, the observations highlight a departure from the simplistic boundaries proposed by the two hypotheses. While an increase in Rcyc correlates with a decrease in $RT_{in\ ms}$, suggesting a faster detection process as the amount of information per unit period increases ($p < 0.01$), the observations are more complex than the two hypotheses would suggest.

The small magnitude of decrease in $RT_{in\ ms}$ with increased Rcyc appears to be supported by the duration hypothesis; however, the influence of information constraints cannot be discounted. This is evidenced by the non-linear trajectory of $RT_{in\ ms}$ across different pattern complexities, suggesting that while the auditory system may prioritize integrate information by a set of time window, it also dynamically integrates item-wise features. Such integration indicates a more adaptable and nuanced information processing than what the duration hypothesis would predict, where cognitive performance is expected to be optimal in comprehending the sensory signal and free from the dominance of temporal limitations.

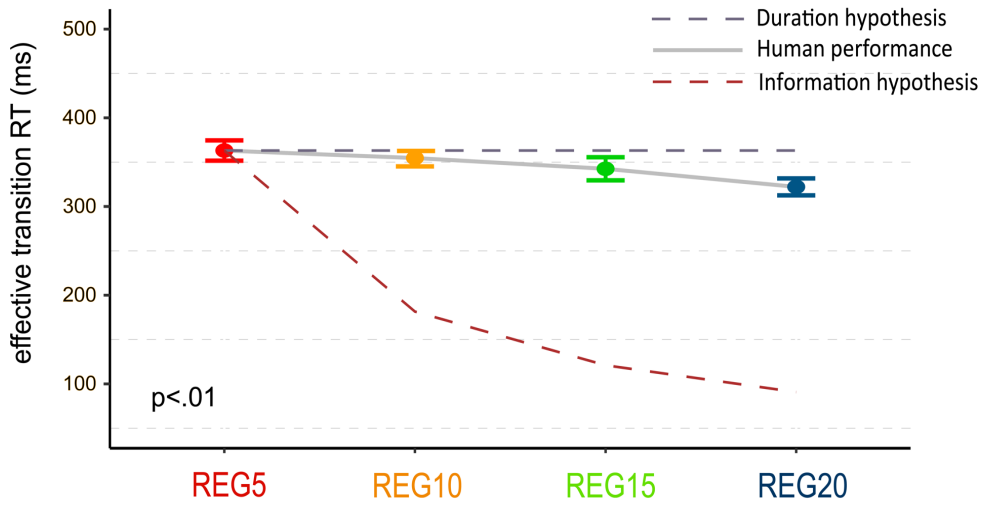

Figure 2.8. Response times fall between the boundaries of two hypotheses. Hypothesis 1 postulates that the memory buffer operates within a fixed time window, processing an auditory pattern regardless of its informational complexity; this is visually represented by the grey dashed line in the figures. In contrast, Hypothesis 2 conjectures that the buffer encodes information adaptively, aligning with the quantity of items as postulated by IdyOM, which is symbolised by the red dashed line. Statistical analysis reveals that human response times ($RT_{in\ ms}$) differ significantly across each experimental condition (rmANOVA, $p < 0.01$). Despite a general trend where $RT_{in\ ms}$ decrease with the increase in Rcyc, there is a pronounced tendency towards the predictions of the duration hypothesis.

2.3. Experiment 2. How is Pattern Detection Performance Affected by Increasing Pattern Duration

Experiment 1 examined the capability of human listeners to detect rapidly presented auditory sequences, with the duration of these patterns fixed at 500 milliseconds. In Experiment 2, the focus was shifted to explore the influence of extended pattern durations

on detection abilities while maintaining constant pattern complexity. To this end, two Rcyc conditions were selected from Experiment 1 - specifically REG10 and REG20 - and measured participants' detection performance at two different pattern durations: the initial 500 milliseconds and a prolonged duration of 1500 milliseconds.

Although 1500ms is relatively slower, it still falls within the echoic memory temporal boundary that has been reported by previous literatures. Therefore, the hypothesis posits that if detection is indeed affected by pattern complexity, a replication of the findings from Experiment 1 within pattern conditions is anticipated.

2.3.1. Methods

2.3.1.1. Participants

105 participants were recruited through Prolific (www.prolific.co) and completed this experiment. Of these, data from 21 participants were rejected due to reports of a noisy environment (see " data pre-processing criteria " below). Data from 10 participants were rejected due to failure to respond to STEP trials or because responses to STEP trials were too slow. In total, 74 participants (26 females; average age 25 ± 4.56 years) were included in the following analysis. In addition, 32 participants did not proceed to the main task due to not passing the pre-determined performance threshold in the practice task, meanwhile, about 27% of participants who initially accessed the experiment but did not pass the headphone screen and therefore did not proceed further (Milne et al., 2020).

2.3.1.2. Stimuli

Four stimulus conditions were used, with each presented in a separate block. Rcyc (REG10 vs REG20) and pattern duration (500 ms vs 1500 ms) were orthogonalised, while tone duration was fixed at 25 ms. Silent gap durations of 25 ms (RANDREG10, presentation rate of 20 Hz) and 0 ms (RANDREG20, presentation rate of 40 Hz) were used to achieve the set pattern duration for the 500 ms condition (Cyc500), while gap durations of 125 ms (RANDREG10, presentation rate of 6.67 Hz) and 50 ms (RANDREG20, presentation rate of 13.33 Hz) were used for the 1500 ms (Cyc1500) conditions. The stimulus set also included CONT and STEP trials as previously described.

2.3.1.3. Procedure

The procedure was similar to that described in Experiment 1. In the main experiment, four test blocks were delivered, each corresponding to one of the conditions mentioned above. Each block lasted 5-7 minutes and included 50 trials (20 RANDREG, 20 RAND, 5 CON, and 5 STEP).

2.3.2. Results

The evaluation of STEP detection efficacy demonstrated stable performance over the course of testing blocks, as indicated by the statistical parameters [$F(3, 219) = .435$, $\eta 2 = .001$, $p = .728$]. Analysis of the d prime data, presented in **Figure 2.9**, evidenced good performance in all experimental conditions. Repeated measures ANOVA, incorporating Rcyc (RANDREG10 vs RANDREG20) and pattern duration (Cyc500 vs Cyc1500) as factors, revealed a pronounced effect of duration on response times [$F(1,73) = 15.17$, $\eta 2 = .052$, $p < 0.001$], and a significant interaction effect between Rcyc and duration [$F(1,73) = 5.45$, $\eta 2 = .956$, $p = 0.022$]. Post hoc comparisons indicated no significant differences in d prime between the RANDREG10 and RANDREG20 conditions within the shorter (500ms) or longer (1500ms) durations. The detected interaction was largely attributable to a slight yet statistically significant improvement in the RANDREG10_Cyc1500 condition as opposed to RANDREG10_Cyc500 ($p<0.05$), underscoring the benefits of slower pace that allows listeners to consciously track the sequence.

The primary analyses in this experiment were firstly focused on the Cyc500 condition to replicate and validate the results of Experiment 1, as depicted in **Figure 2.10**. These analyses confirmed a significant difference in RTs between the Rcyc conditions [$F(1,73) = 805.12$, $\eta 2 = .649$, $p < .001$], with REG20 demonstrating a reduced 40 ms compared to REG10, a finding that echoes the RT patterns observed in Experiment 1, where a mean RT difference of approximately 33 ms was observed.

Subsequent examinations questioned whether similar RT patterns would emerge under the extended pattern duration (Cyc1500). To this end, the RTs between REG10 and REG20 was compared. The findings indicated no significant differences ($p=.433$), suggesting that $RT_{RANDREG10} = RT_{RANDREG20}$, this is equivalent to a relationship of $2RT_{RANDREG10(\text{number of tones})} = RT_{RANDREG20(\text{number of tones})}$ when $RT_{\text{in ms}}$ were linearly transformed as the number of tones. This observation is congruent with the duration constraints hypothesis (Hypothesis 1, see **Figure 2.5**B). Comparing to RT variations in experiment 1 where a decreasing $RT_{\text{in ms}}$ over the increasing of Rcyc was observed, the results of slower sound sequences appear to suggest that the integration process of the memory buffer adheres to a fixed time window.

To compare $RT_{in\ ms}$ across pattern durations, it is more intuitive to convert the $RT_{in\ ms}$ into the $RT_{number\ of\ tones}$. This allows to directly visualise how the amount of information required for pattern detection changes over the increasing of pattern duration. **Figure 2.11** illustrates the $RT_{number\ of\ tones}$ distribution in experiment 2.

Repeated measures modelled on $RT_{number\ of\ tones}$ reveals a significant impact of pattern duration [$F(1,73) = 30.79$, $\eta 2 = .09$, $p < .001$], this indicates that participants processed a greater number of tones when the pattern was presented more slowly, consistent across Rcyc conditions. This finding also aligns with the dynamics illustrated by the PPM decay modelling, which suggests that temporal decay contributes to prolonged detection times (**Figure 2.4**).

Moreover, a significant interaction was observed [$F(1,73) = 21.96$, $\eta 2 = .049$, $p < .001$] between Rcyc and pattern duration, (see **Figure 2.11**B) suggesting that the increase in pattern duration exerts a more pronounced effect on RTs for detecting RANDREG20 compared to detect RANDREG10. These findings suggest a potential shift in the cognitive mechanisms responsible for auditory sensory integration, possibly leaning towards reliance on a fixed temporal window when confronted with an interplay of extended durations and heightened complexity in the presented information.

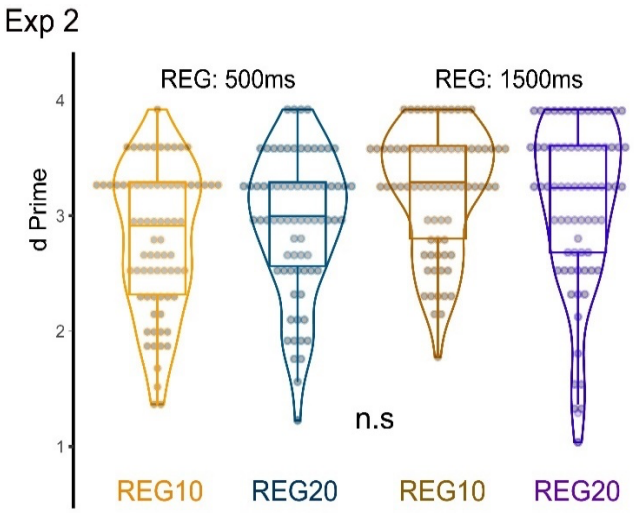

Figure 2.9. d Prime distribution from experiment 2. Most participants present high sensitivity to the pattern emergence, and no significant differences were observed across conditions.

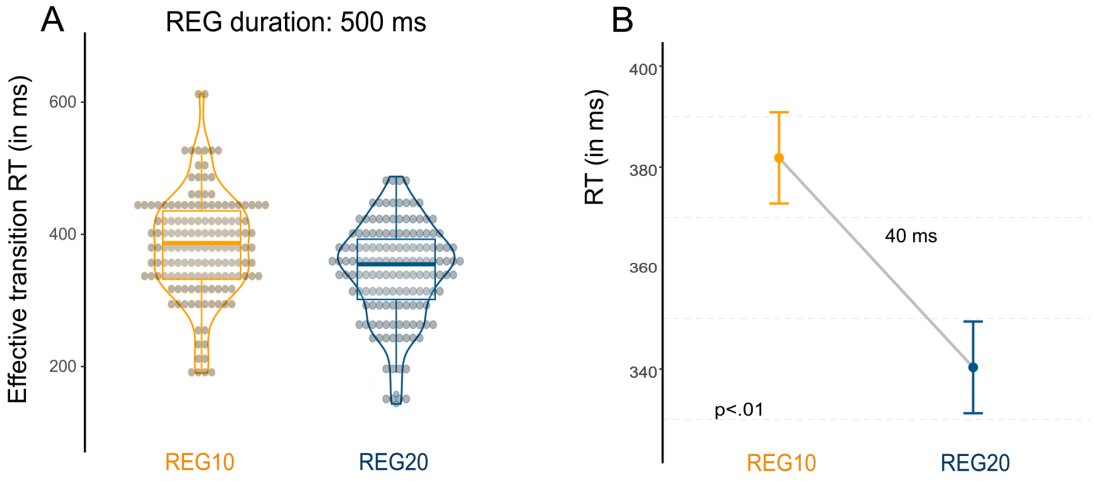

Figure 2.10. The results from experiment 1 were reproduced in experiment 2. (A) Distribution of individual RTs (effective transition) in REG10 and REG20 of fast pattern duration (500ms). (B) The mean and standard error were plotted for each condition. Despite the distinct paradigm and a different group of participants, a consistent response time difference of approximately 40 ms was observed when the Rcyc was augmented to 20.

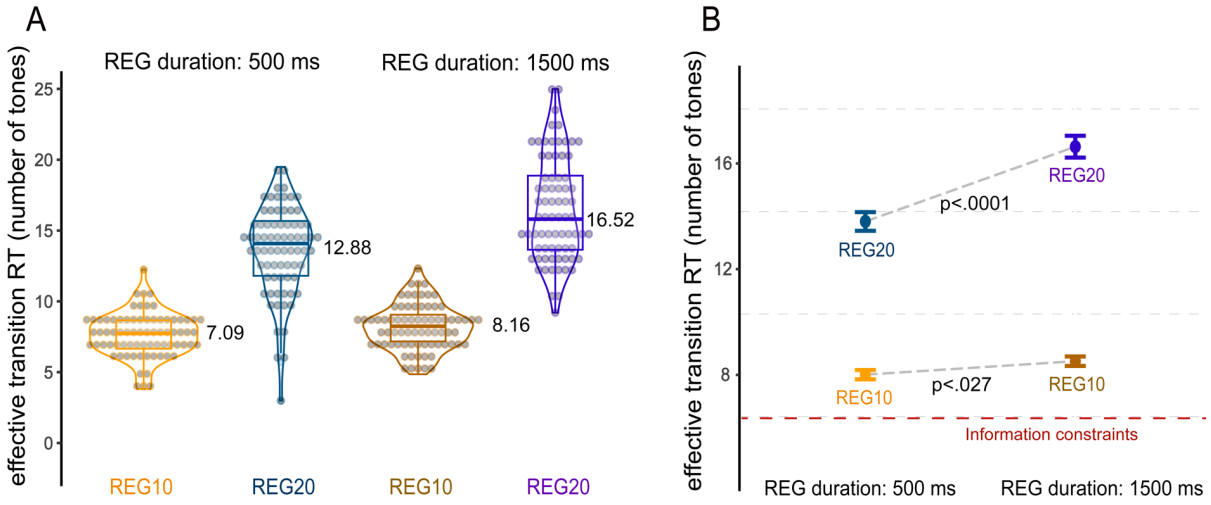

Figure 2.11. Individual distribution of response time across all conditions in Experiment 2. To enable comparison between different pattern duration conditions, The RTs (ms) were converted into their equivalent number of tones to facilitate comparison across pattern duration conditions. (A)

$RT_{number\ of\ tones}$ distribution of all conditions in experiment 2. (B) More number of information was needed in both Rcyc conditions as condition change from Cyc500 to Cyc1500 and the interaction between pattern duration and Rcyc can be identified. Meanwhile, a relationship of $2RT_{RANDREG10(number\ of\ tones)} = RT_{RANDREG20(number\ of\ tones)}$ is observed in pattern duration of 1500ms (Cyc1500), suggests the performance is aligning with duration hypothesis.

2.4. Experiment 3

In Experiment 1 and Experiment 2, the stimulus involved systematically varying the duration of a silent gap between fixed tone-pips, with the tone always lasting for 25 ms. However, it is important to note that the presence of this silent gap introduces potential confounding factors that may affect the auditory perception such as auditory streaming (Moore and Gockel, 2012). For example, introducing silent gaps may allow the brain to perceive sequences of sounds as originating from separate sources, rather than perceiving them as a singular auditory entity. This phenomenon casts doubt on the interpretation of the results from Experiment 1 and Experiment 2. The differences in response times observed across various conditions could be due to the perceptual segregation of the sound sequences. This implies that the memory buffer may internally represent and integrate the multiple sound streams in parallel, potentially leading to quicker detection times for sequences perceived as multiple streams compared to those perceived as a single stream. Although the stimuli in this study maintained a consistent temporal pattern within each testing block, it is known that varying temporal cues play a crucial role in segregation, and that a stream's stability can be facilitated by regular temporal patterns within its elements (Moore and Gockel, 2012). Nonetheless, it remains important to design an experiment that directly manipulates the duration of tones to address the effects of the presence of silent gaps.

To address this issue, experiment 3 was designed with the same experimental paradigm as Experiment 2. The only difference was that the pattern duration of the stimulus was manipulated by varying the tone length. It is expected that, the observations from Experiment 3 will consistently align with the findings from Experiment 2.

2.4.1. Methods

2.4.1.1. Participants:

74 participants were recruited through Prolific (www.prolific.co) and completed this experiment. Of these, data from 12 participants were rejected due to reports of a noisy environment (see " data pre-processing criteria " above). Data from 7 participants were rejected due to failure to respond to STEP trials or because responses to STEP trials were too slow. In total, 55 participants (19 females; average age, 26.3 ± 4.91 years) were included in the following analysis. In addition, 9 participants did not proceed to the main task due to not passing the pre-determined performance threshold in the practice task, meanwhile, about 15% of participants who initially accessed the experiment but did not pass the headphone screen and therefore did not proceed further.

2.4.1.2. Stimuli:

Four stimulus conditions were used, with each presented in a separate block. Rcyc (REG10 vs REG20) and pattern duration (500 ms vs 1500 ms) were orthogonalised. Tone duration of 50 ms and 25 ms (for REG10 and REG20, respectively) were used to achieve the set pattern duration for the 500 ms condition (Cyc500), while tone durations of 150 ms and 75 ms were used for the 1500 ms conditions (Cyc1500). The stimulus set also included CONT and STEP trials as previously described.

2.4.1.3. Procedure:

The procedure was the same as to that described in Experiment 2. However, due to a technical issue, the stimuli set comprising 40 trials (20 RAND and 20 RANDREG) was unintentionally presented twice in a random manner across three blocks of Experiment 3 for all participants. Only REG20 with a pattern duration of 500ms remained unaffected by this issue. Consequently, participants in this experiment were exposed to the stimuli twice as much as in the previous two experiments. This increased exposure has the potential to induce fatigue, which may, in turn, influence response times. Furthermore, the repeated exposure to the same set of stimuli introduces unknown variables that could impact the overall results.

Given these circumstances, this study opted not to analyse the data recorded from this specific experiment in isolation. Instead, the response times data recorded from the first exposure of 40 unique trials for each subject were selected and analysed as a control data set. Nevertheless, the primary objective of Experiment 3 was to control for the influence of silent gaps between tones, which had been introduced in the preceding two experiments. As the result, the study decided to combine and analyse the data from all three experiments

collectively. This approach enabled to solely assess the between-experiment effects that mainly arise from the manipulation of tone length.

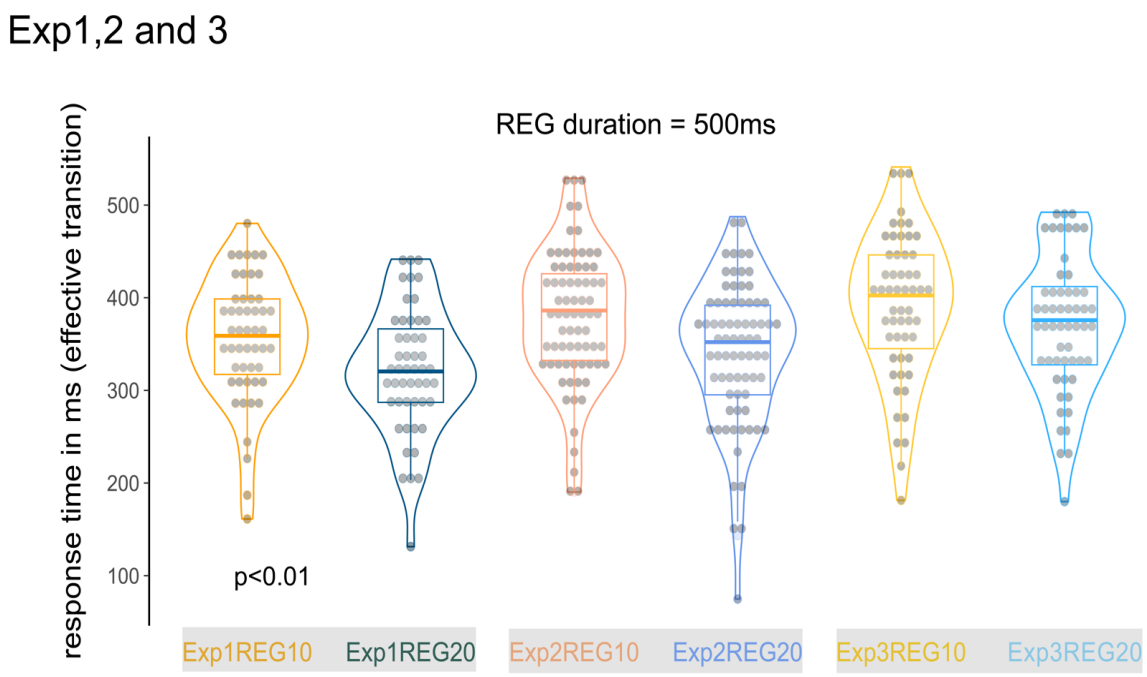

Figure 2.12. Individual $RT_{in\ ms}$ were measured in conditions REG10 and REG20 with a pattern duration of 500ms (Cyc500) in experiments 1, 2, and 3. A significant differences in RTs ($p=0.0018$, repeated measures ANOVA) was observed in all pairs of Rcyc conditions, regardless of the experiment. No interaction of experiment and Rcyc was seen ($p=0.075$).

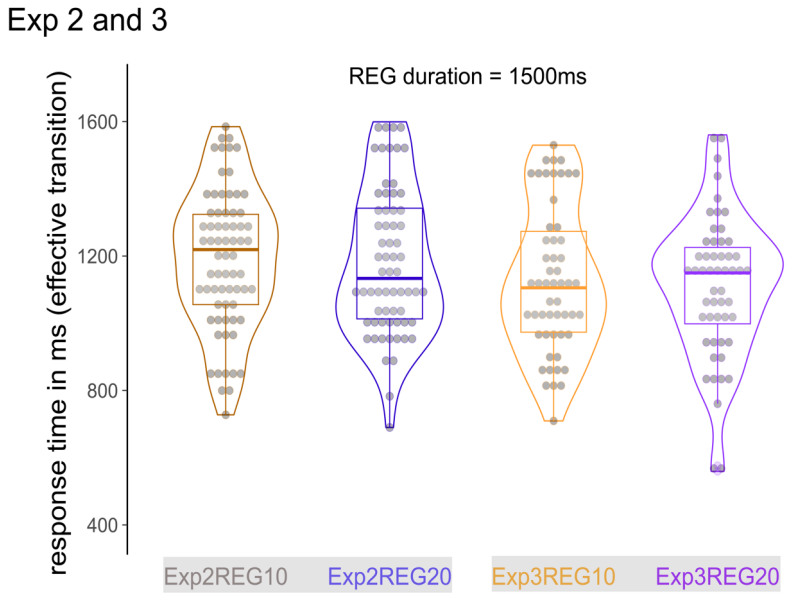

Figure 2.13. Individual $RT_{in\ ms}$ measured in condition REG10 and REG20 with pattern duration of 1500ms (Cyc1500) from experiment 2 and experiment 3. Repeated measures ANOVA suggests no interaction of experiment and Rcyc was seen (p=0.1).

2.4.2. Results

Initially, the detection performance of STEP in Experiment 3 was assessed. The analysis revealed stable performance outcomes [F(3, 162) = 1.94, η2 = .008, p = .12], suggesting participants maintained consistent engagement throughout the experiment. I followed this by conducting a targeted analysis on the Rcyc conditions (REG10 and REG20) within the fast pattern (Cyc500) across all three experiments. This analysis involved a repeated measures ANOVA with Rcyc (REG10 vs. REG20) as the within-subject factor and the experiment number as the between-subject factor for RTs.

The results (**Figure 2.12**) showed a pronounced effect of Rcyc [F(1, 176) = 790.027, η2 = .818, p < .001] and significant variation between experiments [F(2, 176) = 6.84, η2 = .072, p = .001]. Importantly, there was no significant interaction between Rcyc and experiment [F(2, 176) = 2.635, η2 = .029, p = .075], implying that tone length manipulation did not introduce effects on auditory processing while Rcyc varies, as indicated by response times. Post hoc pairwise comparisons revealed significant differences between experiment 1 and 3 (p = .001). That a faster $RT_{in\ ms}$ was seen in experiment 1, compared to experiment 3. This might be due to the increased task time caused by technical issues, which led to

participants feeling fatigued. However, no significant $RT_{in\ ms}$ differences between experiment 2 and 3 ($p = .208$), or between experiment 1 and 2 ($p = .114$) were seen. Those findings confirm the consistent influence of Rcyc across different experiments in short pattern conditions (Cyc500), regardless of experimental conditions and group variability variations.

Furthering the initial analysis, the slow pattern condition (Cyc1500) tested in Experiments 2 and 3 were examined (**Figure 2.13**). The repeated measures ANOVA, with Rcyc (REG10 vs. REG20) as the within-subject factor and experiment as the between-subject factor, showed no effects of Rcyc [$F(1,127) = 0.07$, $\eta2 = .001$, $p = .792$]. In other words, the $RT_{in\ ms}$ between RANDREG10 and RANDREG20 are statistically indistinguishable in both experiments. Different from the Cyc500 pattern conditions, these results seem to align with Hypothesis 1, suggesting the memory buffer which support the pattern detection incorporates information with a fixed temporal window, regardless of pattern complexity.

Significant effects of the experiment [$F(1, 127) = 6.32$, $\eta2 = .047$, $p = .0132$] were found in the slow condition, indicating shorter $RT_{in\ ms}$ in Experiment 3 compared to Experiment 2. This may be due to the longer tone length in Experiment 3, which extended the time for encoding signal perception and therefore enhanced sensory memory, resulting in faster $RT_{in\ ms}$. Crucially, no significant interaction was found between Rcyc and experiment [$F(1, 127) = 2.74$, $\eta2 = .0065$, $p = .1$], corroborating the results observed in the fast pattern condition (Cyc500).

2.5. Discussion

The brain demonstrates exceptional sensitivity to the emergence of sound patterns, a trait that is measurable even when the listeners are distracted. While a multitude of studies have shed lights on the mechanisms behind this phenomenon (Barascud et al., 2016; Southwell et al., 2017; Herrmann and Johnsrude, 2018; Southwell and Chait, 2018; Zhao et al., 2024), there remains plenty of unanswered questions. For instance, what types of information is the brain monitoring? How does the brain integrate and represent sensory inputs in echoic memory that supports the pattern detection?

This study probed the information integration process of echoic memory system using novel tone-pip sequences. To investigate memory mechanisms, it firstly introduced silent gaps between tone-pips to alter pattern duration and complexity. From this, the confounding impacts that could result from merely manipulating tone length was minimised (Harrison et al., 2020). The study generated these sequences from a predetermined set of 20 frequencies (building elements), ensuring that only the tone-pip order changed within each testing block. By manipulating either the informational complexity (Rcyc) or the pattern

duration parameters of the stimuli across the blocks, it was aimed to assess the brain's efficacy to process auditory stimuli varying across different temporal and informational dimensions. Participants were encouraged to detect patterns and to respond by pressing a button as fast as possible upon identifying a pattern within each trial.

The observations from the series of experiments challenged the simplistic models of information integration in memory as proposed by the two hypotheses (Figure 2.5B), which suggested a more dynamic and adaptive mechanism that responds to both the informational complexity and temporal characteristics of auditory stimuli over time when the sound sequence is relatively fast.

2.5.1. Response Times Reflect Dynamic Information Processing in Auditory Memory

To discern auditory patterns, the brain must initially extract the salient features from sounds that activate the auditory system. This study utilised simple-tone pips with variations solely in frequency. Consequently, it was assumed that the extracted features from the tone-pips can be linearly decomposed into distinct pitches (Santoro et al., 2014). Following this extraction, the auditory information is transiently held within the sensory memory buffer, during which the neurons engage in sophisticated computational processes. Theoretically, to 'know' the sequence starts to repeat, the brain must entail a comparison of incoming auditory inputs against those retained within the memory buffer. Additionally, the brain evaluate whether the current auditory stimulus aligns with any pre-existing sequence or denotes a novel sequence. This evaluation encompasses an assessment of the congruence between the memory's stored templates of auditory sequences and the present auditory observations.

The response time documented in this study was adjusted based on the STEP change detection RTs, which accounts for the duration from the sound's entry into the ear, activation of the auditory cortex, and the basic computational comparison of the current sound against an incoming one until the point where the detected alteration reaches conscious awareness and precipitates a motor response. Once the pattern detection response time is corrected by the STEP RTs, the residual duration predominantly encompasses the period required for information processing. This includes the integration and retention of the auditory sequence in memory, and the computational assessment comparing the retained sequence to the current observation.

The first experiment's results revealed a significant inverse relationship between RTs and pattern complexity when the sound pattern is fast (500ms). This means that increased

informational complexity leads to shorter RTs. This finding contradicts the first hypothesis, which suggests that the brain processes sensory signals at a constant rate, regardless of the amount of information per time unit. Instead, our data indicates a more complex neural processing mechanism, where more information speeds up the perceptual detection.

In terms of predictive coding theory, one interpretation of how the echoic memory functions is that it employs a dynamic, internal statistical model to continuously interpret and predict individual sensory inputs. The model assumes that the brain has already integrated these sensory inputs and categorised them as distinct items (tone-pips in this study) before retained them in the memory. Together, the model is able to retrieve past sound inputs stored in the memory buffer to generate the prediction on the upcoming tone. If an imminent tone aligns with the predictions, the resulting reduction in prediction error—or 'surprise'—speeds up the recognition process, leading to a timely response from the participants. Essentially, patterns with more information per unit time seem to improve the memory buffer's predictive accuracy, resulting in faster detection responses.

It is noteworthy that Experiments 1 and 2 yielded similar results. Both experiments showed a consistent approximate 40ms difference in RTs for detecting RANDREG10 versus RANDREG20 patterns. This consistency was observed across different participant groups and experimental frameworks, reinforcing the idea that the brain does not passively receive sensory information and operate them at a constant pace. Instead, it appears to be actively predicting and adapting to the discrete sensory inputs, demonstrating a sensory integration process with an integration window that is shorter than the pattern duration (500ms), likely as short as the duration of the tone-pips used in this study. This is supported by a recent intracranial research (Norman-Haignere et al., 2022), which suggested that the shortest integration window in the auditory cortex is around 30ms.

2.5.2. Increased Informational Complexity Facilitate Pattern Detection but not Behave Like an Ideal Observer Model

Natural sounds are continuous and do not separate into distinct elements like the tone-pip sequences used in this study. To investigate how listeners process statistical information in evolving auditory signals, the process needs to be simplified. One such simplification is the concept of an initial integration process, which turns continuous auditory input into discrete elements. Therefore, this study employed the Information Dynamics of Music model (IdyOM) (Pearce, 2005) to form the basis of hypothesis 2.

While there is the ongoing debate about whether listeners actually interpret auditory patterns this way (Thiessen, 2017), this study chose IdyOM due to the significant amount of previous researches, including computational, behavioural, and neuroimaging studies (Pearce and Wiggins, 2004, 2006; Pearce et al., 2010; Egermann et al., 2013; Bianco et al., 2020; Di Liberto et al., 2020), which showed that the IdyOM can successfully generalise the prediction of musical sequences in human listeners. Meanwhile, increasing evidence have suggested that the brains may innately segment continuous sound stream into their elementary parts, and this strategy appears to be a fundamental feature for how the brain analyses the sound (Poeppel, 2003; Hickok and Poeppel, 2007; Doelling et al., 2014; Ding et al., 2017).

Although the findings from Experiment 1 provided the evidence for an internal statistical model of tracking tone-pip sequences, yet the variation in RTs could not be fully benchmarked by IdyOM. This contradicts Barascud et al. (2016) which indicated that the dynamics of IdyOM mimics the brain's response to random (RAND) and regular (REG) auditory patterns (**Figure 2.1**). Their observations by MEG suggested the listeners might operate similar integration processes as those hypothesised by the model. Nevertheless, a similar pattern detection task by Barascud et al. (2016) found that the listeners needed approximately 15.5 tones to behaviourally detect the pattern emergence of RAND-REG10 — 1.5 tones slower than the brain responses measured by MEG (**Figure 2.1**). These differences indicated the inherent delay between the brain's response to change and the corresponding belief updates to conscious awareness for action. Particularly, as indicated by **Figure 2.5**A, IdyOM shows a decrease in information content after observing 4 tones, and the MEG responses also begin to diverge at this point (**Figure 2.1**). The extra information required for response decision suggested that the brain may need additional computational time to determine if the accumulated evidence is sufficient for triggering the awareness of ‘pattern detected’. This highlighted the potential need for more nuanced models to capture the complexities of the cognitive process.

Furthermore, the online participants in this study required about 7 tones to detect the REG10 pattern, which is 1.5 tones slower than Barascud and colleagues’ observation. Such difference could stem from the factors affecting online listeners, such as a lack of motivation in an unsupervised context, or the environmental distractions that were not present in Barascud's controlled laboratory setting.

2.5.3. Dynamic Nature of Auditory Memory

In Experiment 2, an additional variable—pattern duration—setting it to 1500ms was introduced. The first experiment revealed an inverse relationship between RTs and pattern complexity. This time, the objective was to determine if either of two hypotheses would hold

true for pattern detection efficacy over this extended duration. Although a 1500ms duration seemingly remains within the proposed temporal span of echoic memory (Winkler and Cowan, 2005), the behavioural results uncovered a notable increase in the $RT_{number\ of\ tones}$ for both Rcyc of REG10 and REG20 conditions in detecting the slower patterns, compared to their quicker counterparts (**Figure 2.11**). Crucially, the inverse relationship between RTs and Rcyc, apparent in the faster patterns of Experiment 1 and 2, did not hold in the slower pattern context. Instead, the variation in RTs with the slow pattern aligned more closely with Hypothesis 1, which anticipates consistent RTs for identical pattern durations. This observation implied that the brain processes information within a fixed temporal frame.

These results resemble the findings of McDermott et al. (2013), which revealed the unique auditory representations for fast and slow sound textures. Their research and model indicated that the brain encodes detailed features when processing short sound textures, which allows for a refined discrimination of temporal complexities. In contrast, for longer sound textures, in order to assist the sound differentiation, the brain seems to prioritize statistical summaries of temporal details over certain time epoch. Additionally, the observed interaction between pattern duration and complexity in the second experiment of this study also align with McDermott et al.'s findings, which involved participants tasked with differentiating 'cocktail party' textures that varies either in sound density or duration of the excerpts (McDermott et al., 2013).

The experiments in this study did not precisely identify the type of representation the brain uses during the pattern detection. However, it is evident that with the lengthier patterns, the brain tends to alter its strategy. In specific, it shifts from tracking and predicting detailed item-wise tonal information via a predictive model-like mechanism, to a rougher but efficient processing mode. This mode manifests as the absorbing chunks of information at a steady temporal pace. These findings suggested that the brain's adaptive applications of varying strategies to interpret auditory streams, particularly when they potentially surpass its memory or computational capacities.

2.6. Does the Brain Process Fast and Slow Sounds Differently

The analysis in this study suggested that the brain possesses distinct modes of integration for monitoring rapid and relatively slower sound sequences. From a biological view, this observation appears to be explainable by the brain's adaptive responses, which have been shaped by environmental influences over the course of evolution.

In nature, the brain tracks fast sounds such as the snap of a twig or urgent bird calls, which are crucial for survival and interaction in various environments. These sounds signal the presence of predators or indicate movement, serving as vital cues for both prey and predators (Owings and Morton, 1998; Gerhardt and Huber, 2002; Shelley and Blumstein, 2005). Alternatively, abrupt sounds are commonly to be linked to the appearance or disappearance of a source with the commencement or conclusion of an action. Those moments can be characterised by sharp fluctuations in sound intensity or shifts in the spectral components of the sound, which the brain quickly detects and interprets, as demonstrated by the auditory onset response (Näätänen and Picton, 1987). Since these sounds are likely related to urgent decisions and reactions, the brain thus engages more neural resources to process the information.

For example, the salience network plays a critical role in the brain's response to auditory stimuli, particularly in detecting and prioritising sudden or novel sounds that often indicate important environmental changes (Kayser et al., 2005). This network helps efficiently allocate cognitive resources, ensuring that the brain focuses on relevant auditory signals (Seeley et al., 2007). It enhances processing and collaborating with other networks like the auditory cortex for detailed analysis (Uddin, 2015) and the central executive network for making responses (Menon and Uddin, 2010). In clinical contexts, abnormal responses to auditory stimuli that affect a listener's ability to detect and respond to abrupt changes, like those observed in schizophrenia, are found to be associated with disruptions in the salience network (Todd et al., 2012). Therefore, the integration of the salience network is likely involved in cognitive processes while monitoring rapid sound sequences, facilitating the brain’s ability to capture and analyse detailed sensory inputs.

Contrasting with the transient nature of rapid sounds, slow sounds in the natural environment unfold over more extended periods and are less likely to be associated with threat. For example, environmental sounds such as the continuous murmur of a stream or the gentle rustling of leaves contribute to an auditory backdrop that many species rely on for assessing safety and resource availability (Catchpole and Slater, 2003). To understand the source of those sound, instead of retaining/analysing every detail of them, it is more efficient for the brain to prioritize representing the statistically stable characteristics of the sound (McDermott et al., 2013).

The findings from this study were also supported by the neural evidence provided by Luo and Poeppel (2012). Motivated by question of how the brain integrates speech, the team introduced auditory stimuli composed of three types of 5-second duration segments. Each segment was created by combining individual frequency-modulated tones. These tones had average durations of 25 ms (phonemic rate), 80 ms, and 200 ms (syllabic rate). Participants were instructed to passively listen to the stimuli while their brain activity was recorded by MEG. The data analysis included calculating cross-trial phase and power

coherence. This was done to evaluate the consistency of phase and power patterns across multiple presentations, with a particular focus on the theta (200ms), alpha (80ms), and low gamma (25ms) frequency bands. Their results indicated that stimuli with temporal structures matching natural speech processing scales (approximately 25 ms and 200 ms) elicited reliable phase tracking at corresponding oscillatory frequencies. The theta band response showed right lateralisation, which suggested a hemispheric preference for processing certain temporal scales. In contrast, stimuli with non-matching temporal structures (80ms) did not exhibit phase tracking. These findings demonstrated the auditory system's sensitivity to particular temporal scales, potentially explaining the different modes of integration processes observed in this study (Luo and Poeppel, 2012).

2.7. Conclusion and Future Direction

In summary, this study provides evidence that analysing response times is an effective method for investigating the integration processes of auditory memory. Furthermore, the introduction of silent gaps between tone-pips can be utilised to ensure a fixed duration of low-level sensory encoding. While response times collected in behavioural experiments may be subject to variability due to a range of factors—particularly as this study was conducted online, increasing the likelihood of uncontrolled variables from the unsupervised setting. Nevertheless, this study yielded significant insights into the monitoring strategies humans employ for sound patterns that change in temporal and informational dimensions.

To enhance the understanding of the neural mechanisms involved in auditory pattern monitoring, future research could employ objective methodologies such as electroencephalography (EEG). This approach may illuminate the neural correlates associated with the timing and propagation of brain activity during the processing of sound sequences. For example, the analysis of phase coherence could reveal how the brain synchronises with sensory inputs that vary in Rcyc. Additionally, the event-related potentials evoked by these sound patterns could be quantitatively analysed to decode the underlying neural representations, offering insights into how the brain encode REG pattern that consists of different number of tones (items). Follow on that, fMRI might be utilised to provide precise spatial information about the specific brain regions involved in those processes.

Another interesting question to consider is whether the neural network and brain regions responsible for detecting patterns differ when the duration of these patterns varies, while the number of items within each pattern remains constant. In terms of the results obtained from this study, the hypothesis will be that the brain analyses fast and slow patterns differently. Understanding this difference is beneficial for developing computational models

in the field. However, addressing this hypothesis requires highly sensitive measurement tools; therefore, the employment of invasive methodologies might be necessary to obtain precise data.

3. Chapter 3: Concurrent Encoding of Precision and Event-evoked Prediction Error in Unfolding Auditory Patterns

3.1. Introduction

The physical rules that govern the environment and impose constraints on its agents result in statistically structured, predictable sensory signals. The brain is hypothesised to have developed the capacity to rapidly detect and track the regularities within these signals (de Lange et al., 2018; Press et al., 2020). This ability plays a crucial role in the comprehension of our surroundings, facilitating efficient recognition and processing of incoming information, to empower us to respond rapidly and adaptively to changing circumstances.

The auditory system, in particular, has demonstrated remarkable tuning to regularities across various time scales and dimensions (Bendixen, 2014; Heilbron & Chait, 2018; Carbajal & Malmierca, 2018; Asokan et al, 2019; Fitzgerald & Todd, 2020). This plays a crucial role in our ability to understand spoken language (Arnal and Giraud, 2012), appreciate the nuances of musical compositions (Koelsch et al., 2019) and make sense of the complex soundscape that surrounds us. However, core questions regarding the mechanisms through which regularity is discovered and tracked remain unclear. In particular, pivotal issues revolve around whether the brain chooses to prioritise or suppress predictable sensory signals (Press et al, 2020).

Barascud et al. (2016); *see also* (Sohoglu and Chait, 2016; Southwell et al., 2017; Herrmann and Johnsrude, 2018; Herrmann et al., 2019; Zhao et al., 2024) provided insight into the brain's automatic ability to detect the emergence of predictable acoustic structure by examining low-frequency activity in the M/EEG signal. Using rapidly unfolding (20 Hz) tone-pip sequences that contained transitions from a random (RND) to a regularly repeating pattern (REG), the prior studies observed that a gradual increase in sustained power accompanies the emergence of repeating structures. The timing of the differentiation between REG and RND sequences (3 tones after the first cycle) was consistent with that predicted by an ideal observer model (Pearce, 2005; Harrison et al., 2020), demonstrating statistically efficient processing of structure even when not behaviourally relevant (Barascud et al., 2016).

The sustained response effect is interesting for several reasons: Firstly, it suggests that the brain encodes the inherent state of the stimulus (RND vs REG) rather than merely registering changes in the environment. Secondly, the observed *increase* in sustained power during structure discovery challenges our understanding of how the brain processes and represents predictability. Specifically, it appears to contradict expectations derived from predictive coding frameworks (e.g. Friston, 2005, 2009; Rao & Ballard, 1999), where predictable information is typically associated with *reduced* neural activity, as the brain can efficiently encode and predict upcoming events (de Lange et al., 2018). Barascud et al. (2016) showed that the sustained response, underpinned by activation in the auditory cortex, hippocampus, and inferior frontal gyrus, increases with the predictability of the ongoing stimulus sequence. This prompted the hypothesis that it might reflect the process of tracking the inferred reliability of the unfolding input ('precision'; the accuracy, or conversely the 'expected uncertainty' with which future inputs can be predicted, O'Reilly et al., 2013) whereby predictable sensory streams are associated with heightened sensitivity (see also Zhao et al., 2024).

Several issues need to be addressed for a better interpretation of the sustained response. Firstly, it is important to consider that the effects observed may be specific to the rapid sequences used in Barascud et al. (2016). The behavioural results revealed by the first study also suggests that the brain monitors rapid and slow sound differently. That is to say that evidence from research (e.g. reviewed by (de Lange et al., 2018; Heilbron and Chait, 2018), which focused on slower patterns, might indicate different neural responses. Secondly, it is crucial to determine whether the observed effect primarily reflects a shift in background neural activity or if it also extends to modulations of responses to individual events due to their integration within the structured sequence.

To address these questions, the current study expands upon the stimulus used by Barascud et al, (2016) by introducing silent gaps between successive tones (**Figure 3.1**A and **Figure 3.2**). It was aimed to explore the generality of the sustained-response effects across different temporal scales and provide a clearer understanding of the mechanisms involved in the processing of structured auditory sequences. Additionally, Barascud et al. employed rapid and continuous sound sequences, resulting in overlapping neural responses associated with individual tones. By incorporating silent gaps between tone-pips, it enabled detailed tracking of the neural dynamics corresponding to each sound within the sequences. This approach facilitated a deeper understanding of the underlying neural mechanisms.

3.2. Methods

3.2.1. Experiment 1 - Online Behavioural Study

The behavioural study was designed to probe how the introduction of silent gaps between tones affects explicit pattern detection. It was aimed to pinpoint an optimal gap duration that is sufficiently long to allow us to isolate responses to individual tones, yet brief enough to maintain high-performance levels in pattern detection.

3.2.1.1. Stimuli

Stimuli were sequences of 50-ms tone-pips (gated on and off with 5-ms raised cosine ramps) drawn from a pool of 20 values equally spaced on a logarithmic scale between 222 and 2000 Hz (12% steps). The order in which these tone-pips were successively distributed defined two different sequence types. **RND** sequences consisted of 20 tone-pips (sampled from the full pool) arranged in random order. Each tone-pip occurred equi-probably across the sequence duration. **RNDREG** sequences contained a transition between a RND sequence, and a regularly repeating pattern (REG). REG consisted of 10 different tone-pips, randomly chosen from the full pool on each trial, and repeated in 3 identical cycles. The RND to REG transition always occurred after 30 tone-pips. Opting for this method, as opposed to a variable transition time, ensured a consistent context (in terms of frequency information available) both preceding each transition and across different gap duration conditions. RND and RNDREG sequences were generated anew for each trial and presented equi-probably throughout the experiment. Therefore, the occurrence of a transition in any given trial was unpredictable. The amplitude of each tone pip was normalised to yield an approximately similar perceived loudness (Moore, 2014). Across blocks, the inter-tone-intervals were manipulated to form four conditions (**Figure 3.1** A): **Gap0** (continuous presentation), **Gap100** (a 100 ms gap inserted between tones), **Gap200** (a 200 ms gap inserted between tones), **Gap500** (a 500 ms gap inserted between tones).

Two control stimuli were also included: sequences of contiguous (no silent gap) tone-pips of a fixed frequency (**CONT**) that lasted 4000 ms, and sequences with a step change in frequency partway through the trial (**STEP**: the change always occurred after 2000 ms). These were used to measure individuals' response time to simple acoustic changes and served as 'catch trials' to assess task engagement.

3.2.1.2. Procedure

The experiment was implemented online using the Gorilla Experiment Builder (www.gorilla.sc). Before the main task, participants completed a headphone screening task

(Milne et al., 2020) to ensure they were using appropriate audio equipment. They then received an explanation of the task and completed a practice session. Due to length constraints, the experiment was divided into two parts, performed by two different groups of participants. Experiment 1a contained the Gap0, Gap100 and Gap200 conditions along with the control stimuli (STEP and CONT; see above). Experiment 1b contained the Gap0, Gap100 and Gap500 conditions, along with the control stimuli.

Participants were instructed to respond, by pressing a keyboard button, as soon as possible once they had detected a RNDREG transition or a STEP. To motivate participants to focus on the task, they were given feedback on their accuracy and speed after each trial. A small monetary bonus was given for each correct response (Bianco et al., 2021).

In each experiment, three blocks of 40 trials were delivered. Each block contained the following sequence types: 15 RNDREG, 15 RND, 5 STEP, and 5 CONT. The first block always presented the Gap0 condition. This block lasted 5 minutes. Thereafter, listeners completed the other two blocks (Gap100 and Gap200 in experiment1a, Gap100 and Gap500 in experiment1b) in random order. Starting with Gap0 ensured that all participants experienced the easiest condition first and had adequate opportunity to practice the regularity detection task, reducing the likelihood of frustration and dropout that may occur if participants are immediately faced with the most difficult condition. The main task in experiment 1a lasted about 20 minutes, and that in experiment 1b lasted about 30 minutes.

3.2.1.3. Participant Rejection Criteria

Previous work (Barascud et al., 2016; Bianco et al., 2020) demonstrated that participants are sensitive to the emergence of regularity in RNDREG sequences, exhibiting high sensitivity and rapid detection time (usually responding within two regularity cycles). Due to the online nature of the present experiments and associated reduced control over participants' environments, equipment, and engagement (Bianco et al., 2021), it was important to implement a series of rejection criteria to make sure that data reflect true sequence tracking sensitivity. Therefore, subject data were excluded from the experiment following the below (a-priori determined) criteria:

1) Failure on the Headphone screen: the task introduced by Milne et al. (2020) was used. Participants who did not pass the screening procedure did not proceed to the main experiment.

2) Low performance in the practice run: To ensure participants understand the task, 24 trials with no gap (10 RNDREG, 10 RND, 2 CONT and 2 STEP) were given. Participants did not proceed to the main task if their correct response rate was below 80% in the practice task (see also Bianco et al., 2023). This ensured that those participants who proceeded to

the main experiment could detect the REG transitions, thus allowing us to focus on how performance is affected by increasing the gaps between tones. Our previous experience with similar stimuli in lab settings (see e.g. Barascud et al., 2016; Bianco et al., 2020) suggests that the vast majority of young participants can achieve ceiling performance. We, therefore, reasoned that those online participants who performed below 80% are likely not sufficiently engaged with the task (i.e. distracted, not following instructions, etc).

3) Of those participants who completed the full experiment, the data from those subjects who failed to respond to STEP trials (allowing at most one miss per block) or whose RT to STEP trials fell above 2 STDEV relative to the group mean were rejected. Failure to respond quickly to the (easy) STEP trials indicated low task engagement.

3.2.1.4. Participants

Two participant groups were recruited via the Prolific platform (https://www.prolific.co/).

168 participants took part in experiment 1a. 29% did not proceed to the main task due to failure on the headphone check (this is a similar fail rate to that reported in Milne et al, 2021); 44% did not proceed to the main task due to not passing the threshold of the practice task. This number is much higher than that normally encountered in the lab (see e.g. Bianco et al, 2020 for a similar task) and likely reflects variable engagement by online participants. Data from 5% of subjects were rejected because their STEP responses in the main task were too slow. Data from a further 6% of participants were lost due to network issues affecting the Gorilla online platform. Data from 29 subjects are included in the analysis below (7 females; average age, 24.3 ± 4.79 years).

94 participants took part in experiment1b. 29% did not proceed to the main task due to failure on the headphone check; 21% did not proceed to the main task due to not passing the threshold of the practice task. Data from 10% of subjects were rejected because their STEP responses in the main task were too slow. Data from a further 11% of participants were lost due to network issues affecting the Gorilla online platform. Data from 27 subjects are included in the analysis below (6 females; average age: 22 ± 4.69 years).

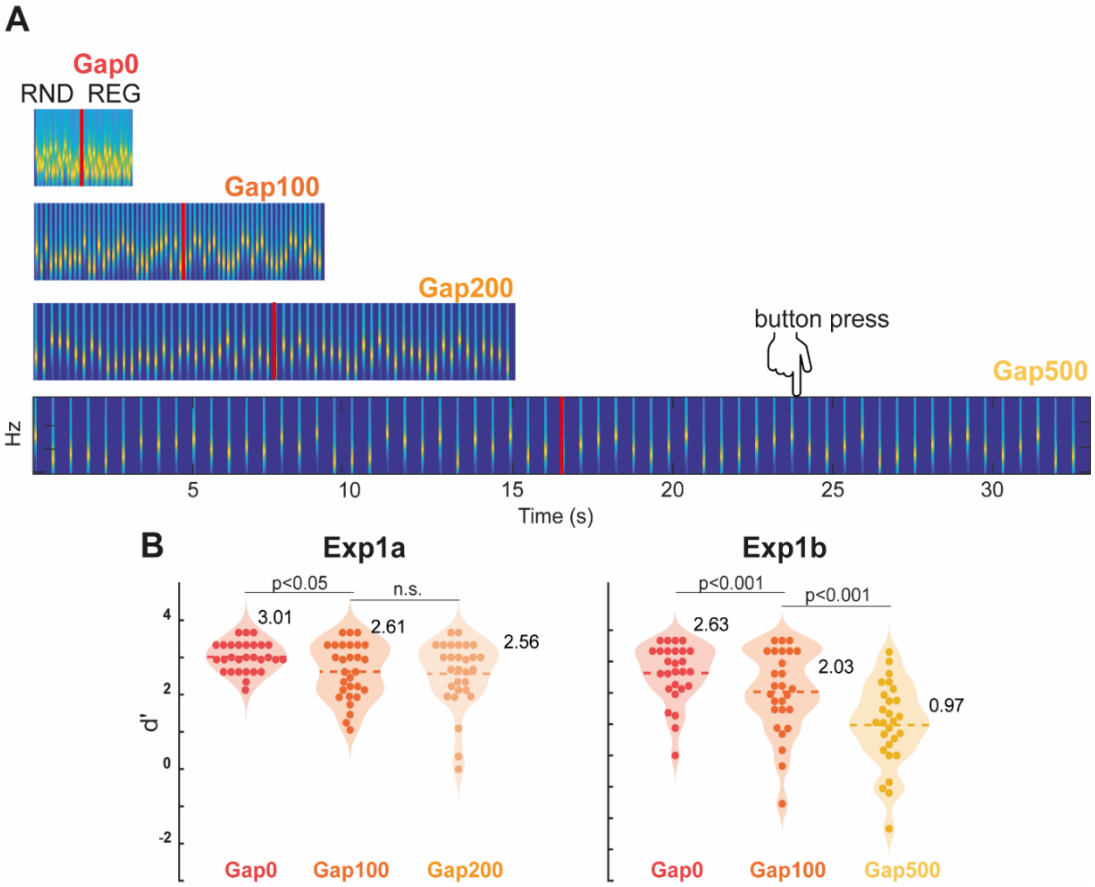

Figure 3.1. Behavioural experiment (A) Examples of the four gap duration stimuli. RNDREG sequences are plotted (the stimulus set also contained 50% no-change RND sequences). Four gap duration conditions are used (0, 100, 200 and 500 ms), resulting in regularity cycles of 500, 1500, 2500 and 5500 ms, respectively. Participants listened to the sound sequences and were instructed to press a keyboard button as soon as they detected the emergence of a REG pattern; indicated with a red line. (B) Behavioural performance. Performance steadily declined with increasing gap duration. Generally good performance (mean d'>2) was seen for the Gap200 condition and it was therefore chosen for the MEG experiment. (Hu et al., 2024)

3.2.2. Experiment 2 - MEG in Naïve Passively Listening Participants

3.2.2.1. Stimuli

Stimuli (Figure 3.2) were generated similarly to those in experiment 1. To reduce the duration of the (passive listening) MEG experiment, the study focused on REG and RND sequences, without transitions. Sensitivity to regularity is investigated by comparing brain responses to the onset of REG and RND sequences. During the initial portion of the sequence (first cycle in REG), responses to the two sequence types should be identical, with differences emerging as soon as the auditory system has discovered that the pattern is repeating. Ideal observer modelling (Barascud et al., 2016; Harrison et al., 2020) suggests that about 3 tones, following the first cycle, are needed for the transition to be statistically detectable. REG sequences were generated by randomly selecting (without replacement) 10 frequencies from the pool and iterating that order to create a regularly repeating pattern. RND sequences consisted of a random succession of 10 tones, newly selected on each trial. All stimuli contained 60 tone-pips. Two timing conditions were used: in ***'fast'*** sequences tone-pips were presented in direct succession (20 Hz rate; 500ms REG cycle duration; 3 s overall sequence duration); in *'slow'* sequences tone-pips were separated by a 200 ms silent gap (4 Hz rate; 2500ms REG cycle duration; 15 s overall sequence duration). One hundred instances of each condition were presented. Sequences were generated anew for each trial such that each stimulus was created of the same frequency "building blocks" (random selection of 10 out of 20 frequencies). Condition presentation was fully randomised.

3.2.2.2. Procedure

The experiment was controlled with the Psychophysics Toolbox extension in MATLAB (Kleiner et al., 2007). All auditory stimuli were presented binaurally via tube earphones (EARTONE 3A 10 Ω; Etymotic Research) inserted into the ear canal, with the volume set at a comfortable listening level, adjusted for each participant.

The experiment lasted 40 minutes. Participants listened passively to the stimuli (presented in random order with an ISI jittered between 1400-1800 ms) and engaged in an incidental visual task. The task consisted of landscape images, grouped in triplets (the duration of each image was 5 s, with 2 s ISI between trials during which the screen was blank). Participants were instructed to fixate on a cross in the centre of the screen and press a button whenever the third image was identical to the first image (10% trials). The visual task served as a decoy task for diverting subjects' attention away from the auditory stimuli. Participants were naïve to the nature of the auditory stimuli and encouraged to focus on the visual task. Feedback was displayed at the end of each block. The experimental session was divided into six 12 min blocks. Participants were allowed a short break between blocks but were required to remain still.

3.2.2.3. Participants

23 naïve subjects participated in the study. One participant's data were discarded due to excessive noise in the data. Data from 22 participants (11 females; average age, 25.14 ± 4.61 years) are reported below.

3.2.2.4. Data Recording and Pre-processing

Magnetic signals were recorded using CTF-275 MEG system (axial gradiometers, 274 channels; 30 reference channels; VSM MedTech). The acquisition was continuous, with a sampling rate of 600 Hz. Offline low-pass filtering was applied at 30 Hz (all filtering in this study was performed with a two-pass, Butterworth filter with zero phase shift). All pre-processing and time domain analyses were performed using the fieldtrip toolbox (Oostenveld et al., 2011). To analyse time domain data, the 40 most responsive channels for each subject were selected. This was done by collapsing across all conditions and identifying the M100 component of the onset response (Näätänen and Picton, 1987; Stufflebeam et al., 1998; Näätänen et al., 2011; Gorina-Careta et al., 2021), as a source-sink pair located over the temporal region of each hemisphere. For each subject, the 40 most strongly activated channels at the peak of the M100 (20 in each hemisphere; 10 in each sink/source) were considered to best reflect auditory activity and thus selected for all subsequent time-domain analyses. This procedure served the dual purpose of enhancing the relevant response components and compensating for any channel misalignment between subjects. Next section will introduce the two-domain analysis pipelines in this study.

3.2.2.5. Whole Sequence Analysis

Initially, responses to the entire sequence were assessed. Low-frequency activity is of prime importance as a possible marker of predictability tracking (Barascud et al., 2016; Southwell et al., 2017). Therefore, no high-pass filter was used. Data were segmented into epochs from 200ms before onset to 1000ms post offset (yielding epochs of 4200ms and 16200ms in '*fast*' and '*slow*' conditions, respectively). Epochs containing artefacts were removed (based on variance summary statistics) using Fieldtrip's manual visual artefact rejection function. Around 5% of epochs were removed from each subject (range 0-10%). The remaining epochs were then averaged by condition. To help denoise the data from *'slow'* conditions (low-frequency drift artefacts) denoising source separation (DSS) analysis was applied to maximize reproducibility across trials. (Särelä and Valpola, 2005; de Cheveigné and Simon, 2008; de Cheveigné and Parra, 2014). For each subject, the three most significant components (i.e., the three 'most reproducible' components across trials) were kept and projected back into sensor space.

3.2.2.6. The Single-tone Response Analysis

A subsequent analysis focused on responses to individual tones in REG vs. RND sequences in the *'slow'* sequences. To identify activity associated with individual tone-evoked responses which might be masked by the sustained activity, the raw data were high-pass filtered at 2Hz. Filtered data were then cut into individual tone epochs, from 50 ms before the onset of the tone, to 200 ms post onset. Responses from tones within each cycle were averaged, yielding 6 time series per condition per subject (tones in Cycle#1, Cycle#2, etc.). Time series were baselined based on pre-tone onset activity.

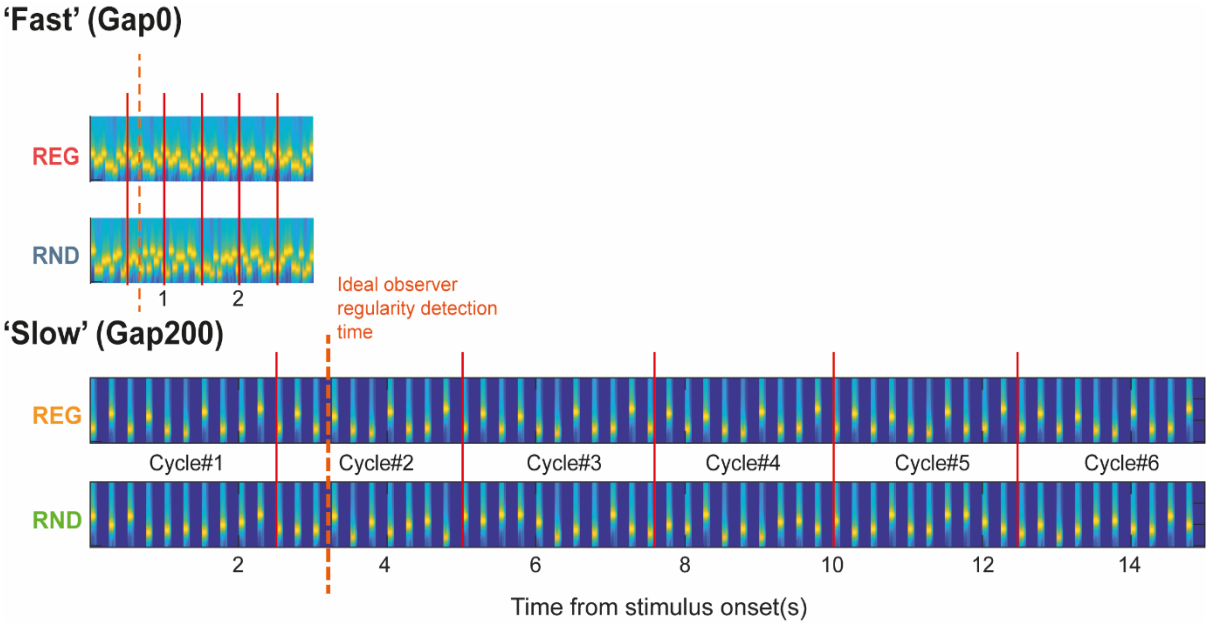

Figure 3.2. Examples of stimuli in the MEG experiment. All stimuli consisted of 60 tones (6 regularity cycles in REG sequences; red lines). 'fast' sequences were 3 s long; 'slow' sequences were 15 s long. Naive participants listened to the sound sequences passively and were instructed to focus on a visual task. If brain responses track the emergence of regularity, responses REG and RND sequences should be differentiated following cycle#1. Ideal observer REG detection latency (~3 tones into the 2nd cycle, e.g. Barascud et al, 2016) is indicated with a dashed line. (Hu et al., 2024)

3.2.2.7. Statistical Analysis

The time domain data are summarised as root-mean square (RMS) across the 40 selected channels for each subject (see above). The RMS is a useful summary signal,

reflecting the instantaneous power of the neural response irrespective of its polarity. Group RMS (RMS of individual subject RMSs) is plotted; statistical analysis was always performed across subjects.

To evaluate differences between conditions (RND vs REG), the RMS differences at each time point were computed for each participant, and a bootstrap re-sampling (Efron and Tibshirani, 1998) was applied (10000 iterations) on the entire epoch. Significance was inferred by inspecting the proportion of bootstrap iterations that fell above or below zero, here p=0.01 was used as a threshold.

3.2.2.8. Source Analysis

To estimate the brain sources that underly the observed time domain effects at the sensor level, source reconstruction using the standard approach implemented in SPM12 (Litvak and Friston, 2008; López et al., 2014; Bartha-Doering et al., 2015) was performed. Sensor-level data were converted from Fieldtrip to SPM. By using 3 fiducial marker locations, the data were co-registered to a generic 8196-vertex inverse-normalised canonical mesh warped to match the SPM's template head model based on the MNI brain (Ashburner and Friston, 2005). This had the advantage of providing a one-to-one mapping between the individual's source-space and the template space, facilitating group analyses (Litvak and Friston, 2008). The forward model was solved with a single shell forward head model for all subjects. Source reconstruction was performed using the multiple sparse priors (MSP) model that assumes that activity can be expressed in multiple patches or covariance components, each of which has an associated hyperparameter (Litvak and Friston, 2008; López et al., 2014; Bartha-Doering et al., 2015). These were optimised with greedy search (GS) technique (Litvak and Friston, 2008) by iterating over successive partitions of multiple sparse priors to find the set yielding the best fit (here a total of 512 total dipoles was specified). The MSP model was used to identify distributed sources of brain activity, hence the 2 conditions (RND and REG) were inverted together.

This study was interested in capturing the sources underlying two aspects of the data:

1. The discovery of regularity (REG vs RND) in the *'fast'* sequence evoked response. The analysis used DSSed data (de Cheveigné and Parra, 2014), with the three most reproducible components projected back into sensor space and used for the inversion. Trials were averaged by condition and the inverse estimates were obtained for the two conditions together using an interval of 300ms between 665 and 965 ms post-stimulus onset. The interval was chosen to coincide with the timing of divergence between the REG and RND conditions as seen in the time domain analysis (Figure 3.3). An attempt was made to analyse the 'slow' sequences (between 3500 and 6000ms post stimulus onset, coinciding with the timing of divergence between REG and RND conditions), but no significant sources were

identified (see supplementary information). This lack of findings can be attributed to several factors, primarily the weaker sustained response effect (see below). Memory constraints probably further exacerbated the issue, resulting in substantial variability across participants when tracking the slow sequences. Importantly, the opposing effects observed for the sustained and tone-evoked responses (see 'results') likely contributed to a net cancellation of effects, making it challenging for the source model to discern meaningful patterns in the 'slow' sequence evoked activity.

2. The effect of regularity (REG vs. RND) on the individual tone responses in *'slow'* sequences. A similar analysis pipeline as that described above was used. This analysis focused on the interval between 5 and 15 s – from the 3rd cycle of the REG until offset, i.e., where the regularity in REG stimuli was well established (theoretically, and, as seen in the time domain data, regularity is discovered partway through the 2nd cycle and well established by the 3rd cycle). The filtered raw signal (2-30 Hz), epoched over 0-200ms post tone onset and averaged across tone presentations, was used for the inversion. The interval was chosen to coincide with the largest possible time window post tone onset to allow the algorithm to encompass all brain sources responsible for generating the response (Henson et al., 2011).

After inversion, source activity for each condition was projected to a three-dimensional source space and smoothed [12-mm full width at half maximum (FWHM) Gaussian smoothing kernel] to create Neuroimaging Informatics Technology Initiative (NIfTI) images of source activity for each subject. At the second level of statistical analysis, the two conditions (REG vs RND) were modelled with the within-subject factor Regularity (REG / RND). Statistical maps of the contrast were thresholded at a level of $p = 0.05$ uncorrected (F contrasts) across the whole-brain volume. Relevant brain regions were identified using the AAL3 toolbox (https://www.oxcns.org/aal3.html).

3.3. Results

3.3.1. Behavioural Performance Reveals Good Sensitivity to Regularity Even Following the Introduction of Silent Gaps between Tones.

This study tested how pattern detection ability is affected by the introduction of a silent gap of increasing length between successive tone pips. Figure 3.1B shows performance (quantified as d' sensitivity score) for each condition in experiments 1a and 1b. With increasing gap duration, an overall gradual worsening of performance was observed. A repeated measures ANOVA over the three gap duration conditions in experiment 1a confirmed a main effect of condition [$F(2, 56) = 3.814$, $\eta 2 = .123$, $p = .026$]. Post hoc tests (Bonferroni corrected) indicated a significant difference between Gap0 and Gap100 conditions [$p = .034$] and between Gap0 and Gap200 conditions [$p = .026$]. No difference between Gap100 and Gap200 was seen [$p = 1$]. In general, most participants achieved a d' above 2 in the Gap200 condition, revealing a largely conserved sensitivity even though the duration of the pattern increased five-fold from 500ms in Gap0 to 2500ms in Gap200. Experiment 1b further tested the performance for silence gaps of 500 ms. A repeated measures ANOVA with factor Gap (0, 100, 500 ms) confirmed a main effect of condition [$F(2, 52) = 33.687$, $\eta 2 = .564$, $p < .001$]. Post hoc (Bonferroni corrected) comparisons indicated significantly worse performance in Gap100 [$p=.025$] and Gap500 [$p<.001$] compared to Gap0, and between Gap100 and Gap500 [$p < .001$].

Overall, the pattern of results is consistent with a slow decline in performance for gaps up to 200ms and a steeper drop thereafter. We, therefore, selected the 200ms gap duration for the MEG experiments (in naïve distracted listeners) below.

3.3.2. The Emergence of Regularity is Associated with an Increase in Sustained MEG Activity

The Group RMS (mean of all subjects' RMSs) of the evoked response to the '*fast*' sequences are shown in Figure 3.3A. The brain response presents prototypical onset activity, followed by a subsequent rise to a sustained response that persists until offset. A pronounced offset response is seen about 100 ms after sound cessation. Fluctuations at 20 Hz, reflecting the tone presentation rate, are visible in the sustained portion of the response.

In line with previous observations (Barascud et al., 2016; Southwell et al., 2017; Southwell and Chait, 2018), REG shows an increased sustained response when compared with RND. The timing at which the response to REG diverges from RND is considered to reflect the information required to detect the regularity. A significant difference between conditions emerged after 665 ms, (13 tone-pips, 1.3 cycles). This estimate is consistent with previous modelling work (Barascud et al., 2016; Harrison et al., 2020) which demonstrated that an ideal observer model required 3-4 tones following the first cycle to detect the emergence of regularity.

Figure 3.3B displays the source analysis, applied over a 300 ms interval over which the REG and RND conditions begin to diverge (yellow shading in Figure 3.3A). The activation map (F contrast, REG>RND, p=0.05) demonstrates increased activity in auditory cortex (AC; bilaterally), inferior frontal gyrus (IFG; bilaterally) and hippocampus (HP; Right Hemisphere only). No areas were identified by using the opposite (RND > REG, p=0.05) contrast. Overall, the source data are largely consistent with what was previously shown by Barascud et al. 2016 for similar stimuli, confirming a distributed network spanning auditory, frontal and hippocampal sources which underlies sensitivity to regular patterns.

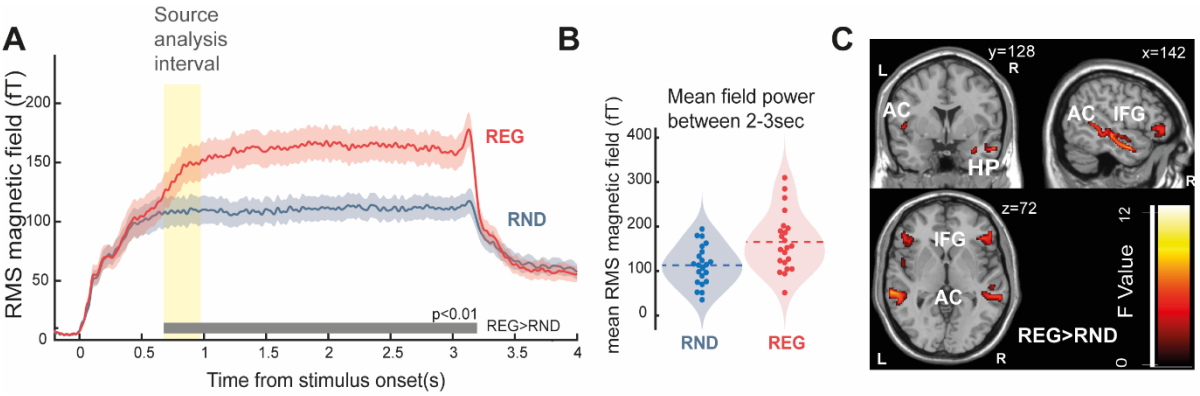

Figure 3.3. MEG response to 'fast' (Gap0) sequences. (A) The full stimulus epoch, from stimulus onset (t = 0s) to offset (t = 3s). The shaded area around the traces indicates the standard error of the mean. The grey horizontal line indicates time intervals where a significant difference is observed between the two conditions (p<0.01). Yellow highlighting indicates the interval (665 ms to 965 ms) used for source analysis in (C). (B) Mean sustained response power computed during the last second of stimulus presentation (2-3 s post-onset) and averaged over trials for each subject in RND and REG conditions. (C) Source analysis. Group SPM F map for the REG > RND contrast during the rising slope of the sustained

response (yellow shaded area in A), thresholded at p = 0.05 (uncorrected). AC: Auditory Cortex; HP: Hippocampus; IFG=Inferior Frontal Gyrus. (Hu et al., 2024)

Responses to the '*slow'* (Gap200) sequences are shown in **Figure 3.4**A. Pronounced fluctuations at 4 Hz, reflecting the tone presentation rate, are clearly visible on top of the sustained response. Similar to what was observed for the '*fast'* sequences, a difference in sustained response emerges between REG and RND when the REG pattern begins to repeat (after 2500ms). This effect is much smaller, however. To separate the sustained response from phasic activity associated with tone-evoked responses, the data were low pass filtered (0-2Hz; **Figure 3.4**B). A significant difference between conditions emerged after 13 tones (3266 ms) consistent with the observations from the *'fast'* sequence above. This suggests that irrespective of the rate at which tones are presented (at least within the range tested here), regularity detection requires a constant amount of information (as measured in number of tones pips). However, it is notable that the sustained difference between REG and RND conditions in the '*slow*' sequences is smaller and rather noisier (e.g. as reflected by the discontinuous significance, see **Figure 3.4**) than in the '*fast*' sequences. A repeated measures ANOVA on the difference between mean sustained response power in REG and RND (as shown in **Figure 3.3**B; **Figure 3.4**B) confirmed a significantly smaller difference between REG and RND in the 'slow' sequences ($F(1, 42) = 18.31$, $\eta 2 = .3036$, $p < .001$).

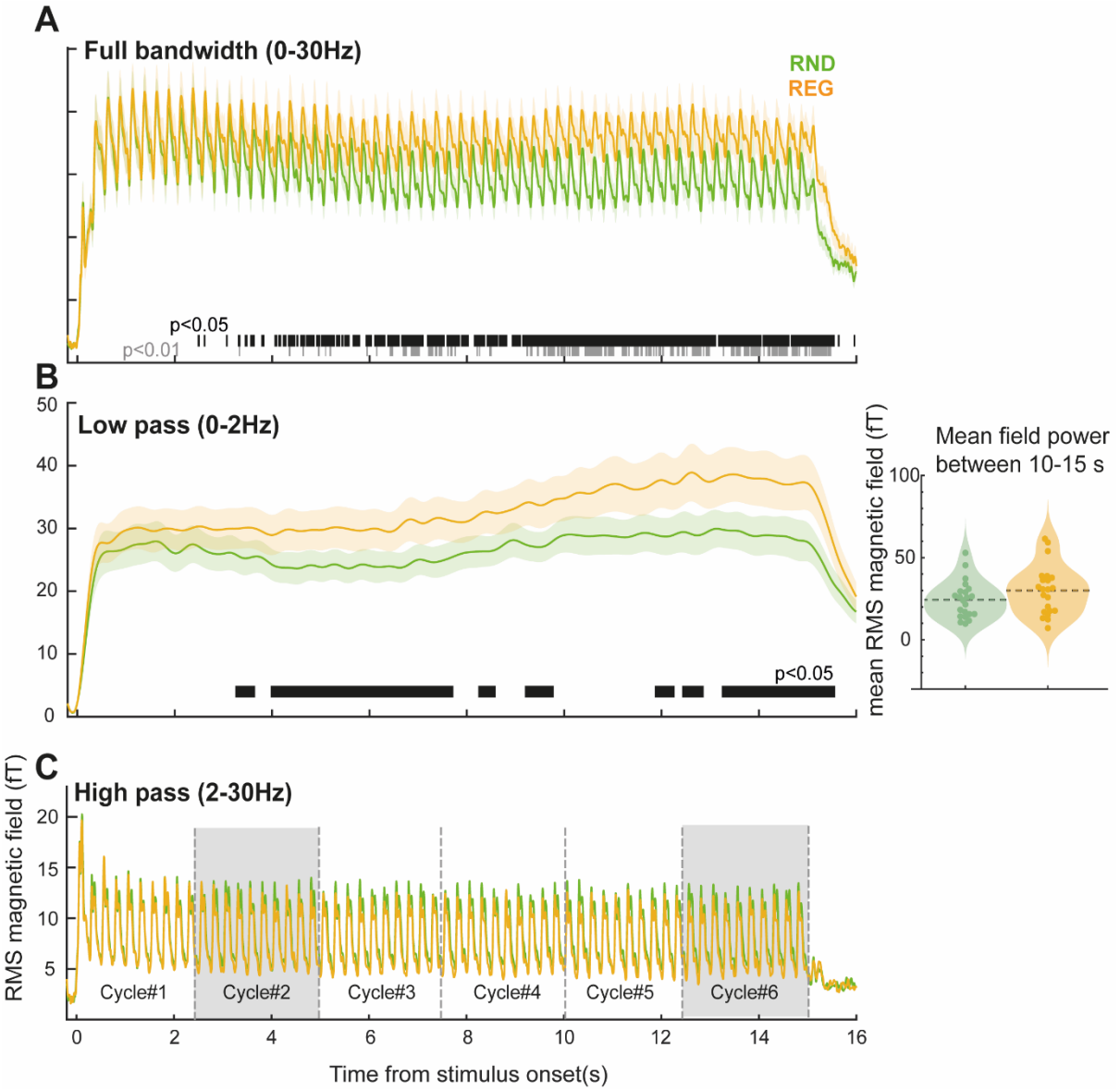

Figure 3.4. MEG response to 'slow' (Gap200) sequences. (A) Wideband; 0-30Hz. The entire stimulus epoch (16s) is plotted. A sustained difference between responses to REG and RND sequences emerges from ~ 3s post-onset. Responses evoked by individual tones (4Hz) are observed throughout the epoch. (B) Low pass filtered responses (0-2Hz) focusing on the slow sustained response activity. The horizontal black and grey lines denote time intervals where a significant difference is observed between conditions ($p < .05$ and $p<.01$, respectively). Mean sustained response power computed between 10-15 s (from the 5th cycle onwards) post-onset for each individual in each condition is shown on the right. (C) High pass filtered activity, with clearly visible responses to individual tones. The 6 REG

cycles analysed in **Figure 3.5** are indicated. Shaded areas are those plotted in **Figure 3.5**B, C. (Hu et al., 2024)

Overall, the MEG results demonstrate that passively elicited brain responses to REG relative to RND sequences are associated with significantly stronger sustained response magnitude, including when pattern durations are long (2500 ms in 'slow' sequences).

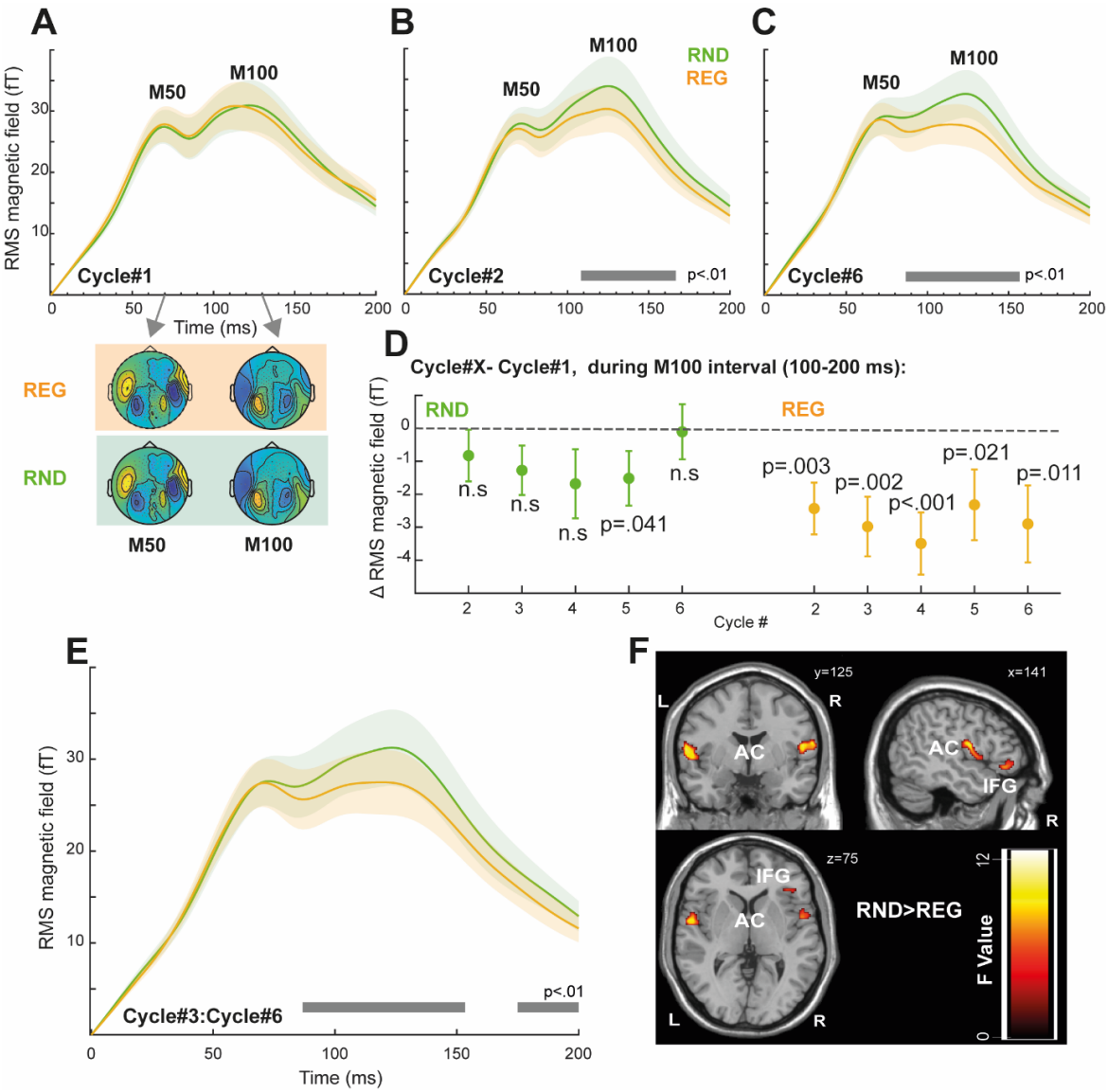

Figure 3.5. Tone evoked responses. (A) Tone-evoked responses averaged over the first 10 tones (0-2.5 s; first cycle) in the RND and REG conditions. Shading around the traces indicates the standard error of the mean. Field maps corresponding to the M50 (60-80 ms) and M100 (130-150 ms) responses are shown below. As expected, no differences are seen because the REG pattern can only be distinguished from RND following the first cycle (once the pattern starts repeating) (B) Tone-evoked responses averaged over tones presented between 2.5 - 5 s in the RND and REG conditions ('Cycle#2). The horizontal grey line indicates time intervals where a significant difference is observed between conditions ($p<.01$) (C) Tone-evoked responses averaged over tones presented between 12.5 - 15 s in the RND and REG conditions ('Cycle#6). (D) Difference from 1st cycle computed (over the M100 time interval; 100-200ms) for each subsequent cycle in REG and RND. Tones presented in REG contexts show consistently reduced activity relative to the 1st cycle. p-values indicate a difference from 0 (one sample t-test). (E) Tone-evoked responses averaged over tones presented during 5-15 s (Cycle#3 to cycle#6). (F) Source analysis results computed from the data in (E). The image is a group SPM F map for the RND > REG contrast, thresholded at $p = 0.05$ (uncorrected). AC: Auditory Cortex; HP: Hippocampus; IFG=Inferior Frontal Gyrus. (Hu et al., 2024)

To focus on phasic activity associated with responses to individual tones, sequence-evoked responses were high pass filtered at 2Hz (**Figure 3.4**C) and tone-centred epochs were extracted (from 50ms pre-tone-onset to 200ms post-tone-onset). The main analysis (**Figure 3.5**), focused on tones presented in each cycle of the REG sequences (see indicated in **Figure 3.4**C; 0-2.5 s; 2.5-5 s; 5-7.5 s; 7.5-10 s; 10-12.5 s; 2.5-15 s), and corresponding tones in RND sequences. As expected, no differences between conditions are seen in the first cycle (cycle#1) (**Figure 3.5**A). In contrast, clear differences between tones presented in REG vs RND contexts are seen in cycle #2 onwards (**Figure 3.5**B; cycle #6 also plotted; **Figure 3.5**C). Critically, REG tones evoke *reduced* responses relative to RND tones. This effect appears to be specific to the latter part of the tone-evoked response: from ~100ms post tone onset, i.e., during the tone-evoked M100 peak.

An additional repeated measures ANOVA on response magnitude (mean power between 100-200ms post tone onset) with regularity (REG vs RND) and tone position in the 2nd – 6th cycles (i.e., from tone #11 to tone #60) as factors revealed a main effect of regularity only ($F(1,21)=4.634$, $\eta2=.181$, $p=.043$), with no effect of tone position ($F(1,49)=1.063$, $\eta2=.048$, $p=.359$) or interaction of the two factors ($F(1,49)=.937$, $\eta2=.043$, $p=.599$). Though

clearly noisy, this tone-by-tone analysis reveals a sustained, stable difference between REG and RND conditions. As a control analysis, a repeated measures ANOVA on the first 10 tones in the sequence (cycle#1) indicated a main effect of tone position ($F(1,21)=9.877$, $\eta2=.32$, $p<.001$) only. Post hoc tests indicated that the responses to the first two tones are significantly different from the third through tenth tones ($p<.01$) in both REG and RND sequences, reflecting increased responses at sequence onset. Neither condition ($F(1,9)=2.647$, $\eta2=.112$, $p=.119$) nor the interaction of condition by tone position ($F(1,9)=.556$, $\eta2=.026$, $p=.832$) were significant. Together, these analyses confirm no difference between REG and RND during the first cycle (cycle#1), with a sustained difference between conditions emerging during the second cycle (cycle#2) onwards.

To further understand whether and how the tone-evoked responses in REG and RND contexts changed over time, the mean evoked field differences between tones presented in the first and subsequent cycles in REG and RND conditions were computed. Because responses to the initial couple of tones (first 2 tones in cycle#1) were affected by onset-response activity, the analysis was focused on the last eight tones of each cycle (cycle#1: tone 3-10; cycle#2: tone 13-20; and so on). The mean tone-evoked response (computed between 100-200 post onset) during cycle#1 was subtracted from that of cycle#2-#6 to understand how the presence of regularity affects tone responses. The data are plotted in **Figure 3.5**D. A repeated measures ANOVA with condition and cycle number as factors yielded a main effect of condition only ($F(1,21)=4.723$, $\eta2=.184$, $p=.041$). No effect of cycle number ($F(4,84)=1.078$, $\eta2=.049$, $p=.373$) or interaction of those two factors ($F(4,84)=1.087$, $\eta2=.049$, $p=.368$) was observed. This indicates a sustained difference between REG and RND conditions, that does not change over time. A one-sample t-test (uncorrected) confirmed that such differences for cycles#2-#6 in the REG condition were below zero, i.e. consistently *reduced* relative to cycle 1. [cycle#2 $t(1,21)=-3.102$,$d=-.661$,$p=.003$; cycle#3 $t(1,21)=-3.288$,$d=-.701$,$p=.002$; cycle#4 $t(1,21)=-3.702$,$d=-.789$,$p<.001$; cycle#5 $t(1,21)=-2.161$,$d=-.461$,$p=.021$; cycle#6 $t(1,21)=-2.478$,$d=-.528$,$p=.011$]. In contrast, the same analysis for RND indicated non-significant effects [cycle#2 $t(1,21)=-1.051$,$d=-.224$,$p=.153$; cycle#3 $t(1,21)=-1.7$,$d=-.363$,$p=.052$; cycle#4 $t(1,21)=-1.604$,$d=-.342$,$p=.062$; cycle#5 $t(1,21)=-1.829$,$d=-.390$,$p=.041$; cycle#6 $t(1,21)=-.125$,$d=-.027$,$p=.451$].

Overall, the tone-evoked analysis demonstrates a consistent difference between tones presented in REG relative to RND contexts, the effect emerges early during the second regularity cycle (i.e. when the regularity has been established) and is manifested as a reduction in responses to REG tones, whilst responses to RND tones remain stable throughout the stimulus period.

Source localisation (see **Figure 3.5**F) for the contrast RND>REG ($p=0.05$) during the tone-evoked response (full epoch – 0-200ms; extracted from the 3rd cycle until sequence

offset; 5-15 s; i.e. after the regularity in REG has been established; see **Figure 3.4**B and **Figure 3.5**E) identified sources in bilateral temporal lobe (superior temporal gyrus, Heschel's gyrus) and bilateral Inferior Frontal Gyrus that underly the time-domain effect. The opposite contrast (REG>RND, p=0.05) yielded no significant activations.

Table 3.1. Summary of MEG source localisation results. MNI coordinates (x,y,z), and F values ($p_{voxel} < 0.05$). Anatomical labelling based on the Harvard-Oxford Cortical Structural Atlas. (Hu et al., 2024)

					MNI Coordinates		
	Region	Side	P-value (peak-level)	F-value	x	y	z
REG-RND	Middle temporal gyrus	Left	0.002	12.42	-56	-28	-10
('*fast*' sequence)	Inferior frontal gyrus	Left	0.026	5.78	-50	34	-4
	Middle temporal gyrus	Right	0.002	12.9	54	-28	-6
	Inferior frontal gyrus	Right	0.024	5.98	46	32	-4
	Hippocampus	Right	0.033	5.22	30	12	-38
RND-REG (tone response extracted from *'slow'* sequence)	Heschl's gyrus/Superior temporal gyrus	Left	0.01	8.09	-60	-8	12
	Inferior frontal gyrus	Left	0.035	5.06	-48	34	-6
	Rolandic operculum	Right	0.035	5.06	52	-4	14
	Inferior frontal gyrus	Right	0.039	4.85	48	28	-8

3.3.3. No Significant Correlation Between Tone-evoked and Sustained-response Effects

To investigate a potential link between the sustained response and tone evoked responses, spearman correlation analysis was conducted on the difference in the tone evoked response (REG-RND; mean power between 100-200ms post tone onset) with a difference in the sustained response (REG-RND; low pass filtered as in **Figure 3.4**B) during Cycle#2 and Cycle#6 across subjects. Both analyses yielded non-significant effects ($p>0.2$).

More complex ridge regression analyses (Bates et al., 2015) was also attempted over single trial data during Cycle#2 and Cycle#6 predicting the tone evoked response with the sustained response and trial number as predictors and subjects as random variable. No significant effects were observed ($p>0.29$).

3.4. Discussion

This study demonstrated that an increased sustained response to regular (REG) compared to random (RND) patterns previously observed in rapid tone sequences (20Hz; 500ms cycle duration), also occurs in slower sequences (4Hz; 2500ms cycle duration). This confirms the auditory brain's remarkable implicit sensitivity to complex patterns. Critically, brain responses evoked by single tones exhibited the opposite effect - lower responses to tones in REG compared to RND sequences. The observation of opposing sustained and evoked response effects reveals parallel processes that shape the representation of unfolding auditory patterns.

3.4.1. Sustained Brain Responses Track Pattern Emergence Even in Slow Sequences

Increased brain responses to predictable, relative to random patterns have previously been documented in many contexts (Barascud et al., 2016; Sohoglu and Chait, 2016; Southwell et al., 2017; Herrmann and Johnsrude, 2018; Herrmann et al., 2019; Zhao et al., 2024). A greater amplitude for REG over RND stimuli is not easily interpretable as a response to physical attributes of the signal. Adaptation, for example, would result in the opposite pattern (Megela and Teyler, 1979; Pérez-González and Malmierca, 2014). Instead, the dynamics of this response, including when it diverges between REG and RND stimuli, suggest that the brain is sensitive to changes in the predictability of sound sequences. On an abstract level, observations regarding how the sustained response is modulated by

sequence predictability suggest it might reflect the coding of precision, or *inferred reliability*, of the incoming sensory information (Barascud et al., 2016; Friston et al., 2017; Heilbron and Chait, 2018; Yon and Frith, 2021; Zhao et al., 2024).

Here it was showed that sustained response effects persist even when sequences are presented at a slower rate (4Hz). Despite the 5-fold increase in pattern duration, the divergence between REG and RND conditions occurred roughly at the same time (3 tones into the second cycle), in *slow* and *fast* sequences, consistent with ideal observer benchmarks (Pearce, 2005; Barascud et al., 2016; Harrison et al., 2020).

It is noteworthy that the sustained response was diminished in the *slow* compared to *fast* sequences. This could be attributed, at least in part, to limitations in human listeners' memory capacity. Indeed, Barascud et al. (2016) observed a reduced sustained response to REG sequences consisting of cycles of 15 tones relative to 10 tones. This was interpreted as indicative of a threshold in encoding patterns that emerges when detecting longer repeating cycles. Similarly, Herrmann et al. (2019) reported reduced sustained responses in older individuals compared to younger participants, hypothesizing that this reduction could stem from age-related decline in tracking regularity patterns (Bianco et al., 2023). To detect the emergence of regularity, the auditory system must presumably maintain and update a statistical model of the auditory input, registering tone repetitions, and decide at which point there is sufficient evidence to indicate a regular pattern. The efficiency of this process relies on the interplay between echoic and short-/long-term memory capacity (Bianco et al., 2020; Harrison et al., 2020). In our study, the introduction of gaps between consecutive tones and the subsequent increase in cycle duration from 500 ms to 2500 ms likely strained short-term memory capacity, leading to less precise memory encoding and therefore overall lower precision for the slow sequences. The behavioural results indeed indicate a decline in pattern detection (Figure 3.1). However, it is crucial to emphasize that despite this decline, the mean performance level remained high, underscoring the largely preserved sequence tracking capacity.

The brain mechanisms underlying the sustained response remain unclear. Source analysis suggests that the amplified response is driven by cortical activation in auditory, IFG and hippocampal sources (see also Barascud et al. (2016). A similar network involving the auditory cortex and IFG has been implicated in the generation of the Mismatch Negativity response (Näätänen et al., 2012) and has been postulated to represent the circuit responsible for maintaining an auditory model and conveying predictions to lower processing levels (Garrido et al., 2009; Heilbron and Chait, 2018).

According to one interpretation, the sustained response might reflect an excitatory processing mechanism, characterised by an increase in gain, potentially via neuromodulation, on units responsible for encoding reliable sensory information (Feldman

and Friston, 2010; Auksztulewicz et al., 2017). In particular, tonic Acetylcholine (ACh) has been shown to be modulated by environmental uncertainty (Dalley et al., 2001; Yu and Dayan, 2005; Bland and schaefer, 2012).

However, this interpretation may be less tenable, as it predicts heightened responses to tones within the REG sequences, which is contrary to our observed findings (see below). Alternatively, the sustained response may indicate an enhancement in the inhibition of neuronal units that convey low information content. This is consistent with prior research, albeit involving simpler stimuli, where an increase in inhibitory activity linked to the presence of predictable information has been documented (Natan et al., 2015, 2017; Schulz et al., 2021; Richter and Gjorgjieva, 2022; Yarden et al., 2022). A specific role for inhibition, instead of excitation, in governing responses to predictable sensory stimuli, is also supported by indirect evidence from dynamic causal modelling (Lecaignard et al., 2022) and behavioural findings: rather than capturing attention, predictable patterns are more easily ignored (Southwell et al., 2017) and are linked to reduced arousal (Milne et al., 2021). It is important to emphasize that M/EEG (or BOLD) do not readily differentiate between inhibitory and excitatory activity. Therefore, further advancement in understanding this phenomenon necessitates focused investigations at the cellular level.

3.4.2. Reduced Responses to Tones in REG Relative to RND Patterns

Introducing temporal gaps between successive pips allowed us to disentangle the neural responses elicited by individual tones. Results revealed a reduction in neural activity in response to tones embedded within regularly repeating relatively to random patterns. This effect appears to be driven by relatively stable responses to tones in random patterns, but declining responses in the REG context. The dynamics of this effect are consistent with a step change in response magnitude during the second cycle (after the regularity has been introduced) that is then fixed for the remainder of the sequence.

Reduced response to REG tones is consistent with predictive coding theories (Rao and Ballard, 1999b; Lee and Mumford, 2003; Friston, 2005, 2009). According to these models, top-down expectations, derived from statistical regularities in the external world, play a crucial role in suppressing anticipated sensory input. This mechanism serves as an efficient neural coding scheme, optimizing the allocation of neural resources and enabling the brain to prioritize the processing of novel or unexpected information, which may hold greater relevance (Olshausen and Field, 1996, 2004; Friston, 2005, 2009; Tang et al., 2018). Empirical support for these predictions, often referred to as 'expectation suppression', has been mounting across sensory modalities, (Baldeweg, 2006; Summerfield et al., 2008;

Alink et al., 2010; Ouden et al., 2010; Todorovic et al., 2011; Kok et al., 2012; Todorovic and Lange, 2012; Barbosa and Kouider, 2018; Heilbron and Chait, 2018). In the auditory domain, Todorovic and de Lange (2012) demonstrated that when tones were expected based on the probability structure of tone transitions, they elicited suppressed auditory activity within a specific time window of 100–200 ms. This suppression was uniquely attributable to the phenomenon of expectation suppression and distinct from adaptation (repetition suppression) effects.

Notably, the effects that was reported manifest within this same time-window (100-200ms; during the M100 phase of the response). Whilst it is difficult to exclude low-level processes such as adaptation, several patterns in the dynamics of the development of these effects suggest that simple adaptation is unlikely to be a main factor. Firstly, the effects require processes that persist for 2500ms (duration of a cycle). Secondly, no gradual reduction in responses to REG tones that builds up over cycles was seen. Rather there is a step change in the second cycle that is then consistent for the remainder of the sequence.

3.4.3. Multiplexed Representation of Sequence Predictability

The challenge faced by sensory systems is to accurately and swiftly represent information to support adaptive behaviour and facilitate interaction with the environment. A fundamental question pertains to whether the brain primarily encodes predictable or novel information (Press et al., 2020). Bayesian cognitive models propose that our predisposition to perceive what we expect enhances the fidelity of our sensory experiences (Wyart et al., 2012; Summerfield and de Lange, 2014; Kaiser et al., 2019). In contrast, cancellation models suggest that our perceptual system prioritises unexpected stimuli, as they carry an informative value (Blakemore et al., 1998; Meyer and Olson, 2011; Richter et al., 2018). In line with these considerations, predictive coding models (Rao and Ballard, 1999; Friston, 2005, 2009) postulate the existence of two functionally distinct subpopulations of neurons within the brain. One encodes the conditional expectations of perceptual causes, while the other encodes prediction error.

Our findings confirm the coexistence of these facets of regularity coding within the MEG signal: the sustained response is consistent with the encoding of the precision of the signal, whereas responses to individual tones appear to correspond to the coding of prediction error, as indicated by the reduced responses to predictable tones. Intriguingly, our results underscore the active involvement of the same neural network, encompassing the auditory cortex and the IFG, in both discovering structural patterns within auditory sequences and dampening responses to anticipated stimuli. However, the spatial resolution

limitations inherent to MEG source analysis prevent definitive conclusions about the precise co-localisation of these neural processes.

Indeed, the question of whether these manifestations stem from a singular process exhibiting differential characteristics in sustained and tone-evoked responses, or if they represent two distinct mechanisms, as proposed in previous works (Rao and Ballard, 1999; Friston, 2005, 2009), emerges as a crucial avenue for future exploration. For example, it is possible that the sustained response reflects activity linked to a tonic inhibitory drive (implementing gain control) onto sensory units, resulting in a diminished evoked response to individual stimuli. Notably, our study did not reveal a correlational relationship between tone-evoked and sustained responses. While this may tentatively suggest no direct linkage between the two mechanisms, it's essential to consider the possibility that this observation could be influenced by the inherent noise in MEG measurements. More nuanced insights will be gleaned with the application of sensitive invasive tools in future investigations.

3.5. Supplementary Information

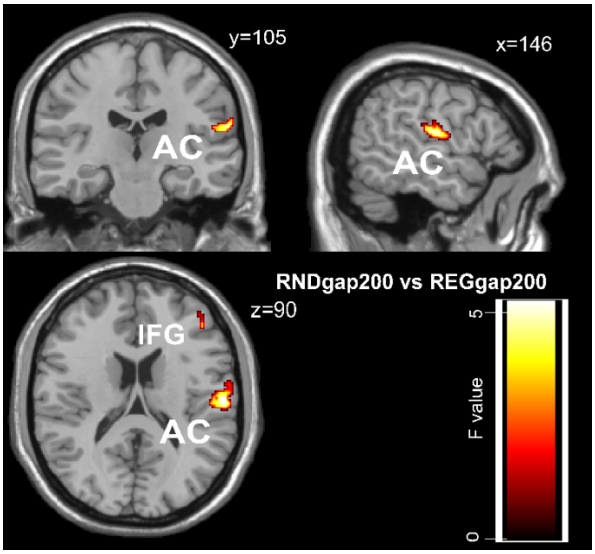

Figure 3.6. The source analysis on the sequence evoked response on 'slow' conditions (between 3500 and 6000ms post stimulus onset, coinciding with the timing of divergence between REG and RND conditions) was conducted by using a p=0.1 and the results confirm the ubiquitous Auditory Cortex (AC) and IFG pattern and consistent with the fact that Right Hemisphere sources are often more strongly activated than Left hemisphere ones. The Hippocampus source is usually weak, so it is not surprising it is not evident here. It appears to be for a more robust source analysis substantially more trials/subjects would be needed.

4. Chapter 4. Whether/which Cognitive Factor Might Account for the Variability in Pattern Detection Performance.

The results from Study 2(see Chapter 3) indicate that listeners' sensitivity (d' score) to slow sound patterns (Gap500) exhibits significant variability; while some participants achieved ceiling performance, others performed at chance level. To explore the underlying causes of this variability in detecting RNDREG transitions (the emergence of a regular pattern), several cognitive factors were considered:

Working/Short-term Memory Ability: Although listeners in the pattern detection task were not explicitly instructed to memorize sounds, recognizing patterns required them to maintain a mental representation of previous signals for comparison with new sounds. Kumar et al. (2014) found that human subjects could learn to identify repetitive noise patterns without direct guidance, demonstrating unsupervised learning through extended auditory exposure. The fMRI and multi-voxel pattern analysis revealed that both the planum temporale and the hippocampus were adept at differentiating between familiar and new acoustic patterns. This underscores the pivotal role of the hippocampus in storing long-term auditory experiences that facilitate pattern recognition (Kumar et al., 2014); similarly, the MEG data reported by Barascud et al. (2016) also suggests the involvement of hippocampus, collaborates in discriminating between regularly repeating and random tone patterns during passive listening (Barascud et al., 2016). Prompted by those insights, this study is investigating whether auditory pattern detection is dependent on working or short-term memory capabilities, functions also linked to hippocampal activity (Kumar et al., 2016; Hauser et al., 2020).

Sustained Attention Ability: Successful performance in the task may require listeners to continuously focus on the unfolding sound stream.

General Vigilance/Task Engagement: This encompasses overall engagement/motivation with the task at hand.

To identify which factors might influence performance, participants completed the pattern detection task (termed as 'Gap100', 'Gap500' task in this study) alongside a battery of cognitive assessments designed to measure working/short-term memory, sustained attention, frequency discrimination, and general task engagement.

This study was structured into two sub-studies:

The first sub-study compared auditory and visual sustained attention tasks to determine which better reveal the individual variability of sustained attention ability.

The second sub-study focused on identifying the cognitive factors that best explain the observed variability in the explicit pattern detection task among human listeners.

4.1. Sub-study 1: Comparison Between Auditory SART and Visual SART

4.2. Introduction

Sustained attention refers to the ability to maintain consistent focus on a particular aspect of a stimulus over an extended period. It is typically assessed through monitoring tasks that require participants to detect a target among irrelevant signals presented sequentially (Esterman and Rothlein, 2019). The Sustained Attention to Response Task (SART) is a sensitive measure for evaluating sustained attention abilities (Robertson et al., 1997). In its visual version, participants monitor a sequence of digits displayed one after another. They must press a button in response to frequent non-target stimuli ('go trials' labelled '1-2, 4-9') and withhold their response for the less frequent target stimulus ('Nogo' trial, number '3'). This setup involves 25 'go' trials and 200 'no go' trials. Importantly, the commission error, which is the failure to withhold the response to 'Nogo' trials, are the most common measure in SART. Accumulating work suggest that commission error rate during the SART coincide with subjective reports of mind wandering (Smallwood et al., 2007, 2008), and the propensity for making errors is correlated with self-reported measures of absent-mindedness (Cheyne et al., 2009).

SART has been widely used to address various clinical populations that have difficulties with sustained attention. For instance, evidence suggests that variations in the dopamine β-hydroxylase (DBH) gene are linked to attention deficits and broader executive function deficits in individuals with ADHD. The study by Greene et al. (2009) explored the relationship between genetic variations in the DBH gene and sustained attention. It specifically examines how the DBH C-1021T polymorphism affects the conversion rate of dopamine to noradrenaline, noting that the T allele is associated with a slower conversion rate than the C allele. This difference may lead to varied levels of cortical dopamine and noradrenaline, influencing neurological processes related to attention. The study involved

200 participants who were genotyped for the DBH C-1021T marker and then completed the Sustained Attention to Response Task (SART). The findings indicate that the DBH genotype significantly influences performance on this task, with individuals carrying more copies of the T allele exhibiting more errors of commission (Nogo hits), indicative of lapses in sustained attention. These errors were correlated with decreased noradrenaline levels, implying that a slower conversion rate may contribute to reduced alertness and a heightened susceptibility to distractions (Greene et al., 2009).

Another related study by Farring and colleagues (2003) investigated the effect of depression on sustained attention while using SART. Their work exposes a mismatch between how people with depression evaluate their cognitive abilities and what objective assessments reveal. The study included 102 UK servicemembers who were divided into "depressed" or "nondepressed" groups based on their Beck Depression Inventory scores. All participants took the SART. The results showed that those with depression made more commission errors than those without. Moreover, the depressed group subjectively reported experiencing more cognitive failures than they objectively performed, indicating a higher perceived level of cognitive impairment (Farrin et al., 2003).

Except for ADHD and depression, evidence suggests schizophrenia was associated with sustained attention deficit revealed by SART. Chan et al. (2009) conducted a study investigating sustained attention deficits across various levels of psychosis proneness using the SART. The study included 199 participants, divided into three groups: 74 individuals diagnosed with schizophrenia, 69 individuals identified as having schizotypal personality features due to high scores on the Schizotypal Personality Questionnaire (SPQ), and 56 healthy control participants. The results revealed that both individuals with schizophrenia and those with schizotypal features performed worse (obtained more commission errors) than the healthy controls (Chan et al., 2009).

Furthermore, the crucial role of sustained attention in self-awareness was also demonstrated among patients with traumatic brain injury (TBI). The research highlights the importance of self-awareness in affecting rehabilitation outcomes. It explains how cognitive function impairments like diminished sustained attention can significantly affect patients' understanding of their own cognitive deficits. Methodologically, the study evaluates TBI patients' self-monitoring capabilities through an online error-monitoring task (visually presented symbols or letters) that requires sustained attention, thereby assessing how these patients perceive and adjust to their actions throughout the duration of this task. The findings present a strong connection between commission errors and both types of online self-awareness—emergent and anticipatory. They highlight that the ability to maintain attention correlates closely with recognizing cognitive failures (O'Keeffe et al., 2007).

Traditionally, SART has relied on visually presented digits or letters/symbols as used by O'Keeffe et al. (2007). However, in this experiment, both visual and auditory version of the task were introduced, where in auditory task, visual digits are replaced by spoken digits. In fact, this experimental design was initially addressed by Seli and colleagues that aimed to assess whether individual differences in sustained attention, as measured by the visual SART, would remain consistent in an auditory format. Within the experiment, participants completed three task blocks across auditory, visual, and combined auditory-visual modalities. Each block consisted of 225 trials displaying digits from 1 to 9 randomly. In the visual SART, these digits appeared on a screen, whereas in the auditory SART, they were spoken through headphones. Participants were required to press a key for all digits except the number 3 (Nogo trials), where they were to withhold their response. Their findings indicated that the auditory SART typically resulted in slower response times and fewer errors than the visual version. Despite these performance discrepancies, strong correlations between the two modalities demonstrated consistent measures of sustained attention (Seli et al., 2012a).

As the result, in this sub-study, the first objective is to replicate their findings (Seli et al., 2012a). Secondly, it is aimed to evaluate the performance differences between these two modalities and examine which task can optimally capture the sustained attention variability.

4.3. Methods

4.3.1. Visual SART

The Gorilla experiment platform was utilised to implement a version of the task. Digits were displayed at the centre of a computer screen in one of five randomly assigned font sizes (48, 72, 94, 100, and 120), corresponding to digit heights between 12 and 29 mm. The main task involved the delivery of 225 single digits. Each digit was shown for 250 ms, followed by a 900 ms mask composed of a 20 mm ring with a diagonal cross in the centre. The presentation was paced at an onset-to-onset interval of 1150 ms. Both the digits and the mask were white, set against a black background. The outcome measures were the percentage of failures on no-go trials.

4.3.2. Auditory SART

The same experimental platform was used for the auditory version of the SART test. Stimuli were presented as single spoken digits, randomly spoken by either a pre-recorded

male or female voice. The main task delivered 225 spoken digits. Each stimulus was presented at an onset-to-onset interval of 1150ms (stimulus + silence) to align with the visual SART. As the duration of each spoken digit varied (0.35 ± 0.0516 s), the silence after the stimulus also varied slightly. To minimize visual distractions, participants were instructed to focus on the fixation cross at the centre of the screen. The outcome measures were the same as those for the visual SART.

4.3.3. Participants

Participants were recruited through Prolific (www.prolific.co). Each participant was presented with two tasks in a shuffled order, and practice was provided before proceeding to the main task. Data from 24 participants (11 females), with an average age of 22.8 ± 3.3, were included in the following analysis.

4.4. Results

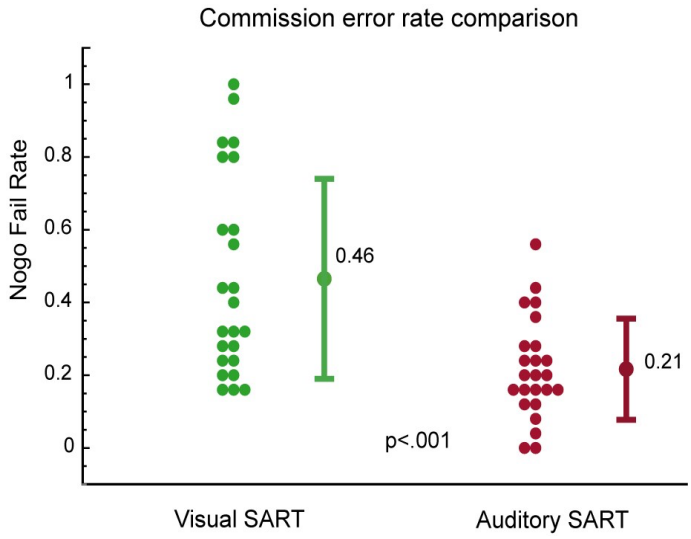

Figure 4.1. Distribution of commission errors (hits on Nogo trials) across all NOGO trials for each of auditory, visual version of Sustained Attention Response Task (SART). Significant differences were seen between modalities (p<.001).

4.4.1. NoGo Commission Errors

Differences (**Figure 4.1**) were seen [t (23) = 4.386, p < .001] between the commission errors (responding to Nogo trial) of auditory SART (mean = 0.21, std = 0.13); and visual SART (mean = 0.46, std = 0.27). Visual SART yielded significantly higher error rate than the auditory SART. However, contradicting to the previous study (Seli et al., 2012a), which they proposed the strong correlation of proportion error across the two versions of the SART, no correlation (spearman) in commission error rate between auditory and visual SART was seen (N=24) in this study. Considering the 'noise' of the data due to the unsupervised environment, this study selectively analysed the sub-group (participants whose commission error rate below 80% in visual SART) as it was assumed that healthy participants who were fully engaged in the task can reach that threshold. Therefore, data from 19 subjects were included. However, no significant correlation of commission error rate between two tasks was observed, despite only the subjects with reasonable commission error rate in visual SART (below 80%) were included [r(18)=.255,p=.277].

4.4.2. Go Trial Response Times

Shorter response time in visual SART with mean of 385ms relative to auditory SART with mean of 627ms were seen [t(23)=-9.193, p <.001]. No correlation in response times across two modalities [r(23)=0.2443,p=0.2487] when all subjects were included; However, significant correlation emerged in those subjects whose commission rate was below 80% in visual SART [r(18)=0.5035,p=0.0297] (**Figure 4.2**).

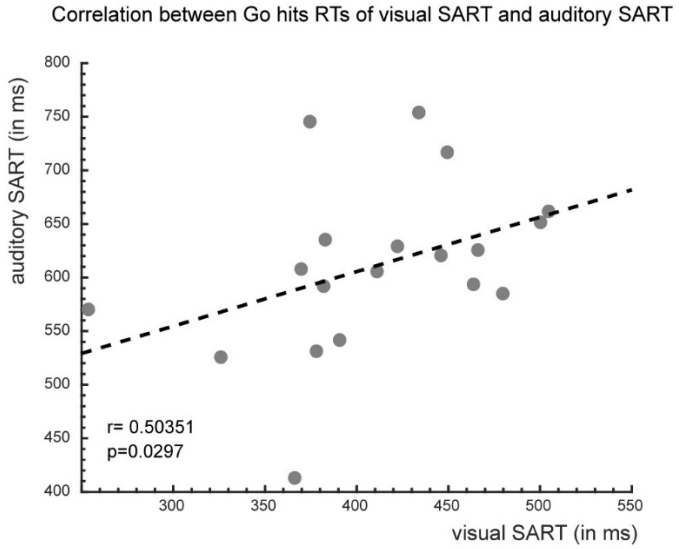

Figure 4.2. Significant correlation was seen across modalities in subjects group whose NoGo commission error rate was below 80% in visual SART (N=19).

4.4.3. NoGo Trial Commission Error Rate and Go Trial Response Times

Interestingly, significant negative correlation was seen between NoGo commission error rate and Go hits response times in visual SART [r(23)=-0.916, p <.001] (all subjects were included), and this observation was not obtained in auditory SART [r(23)=-0.1546, p = 0.4708], although in both tasks, the NoGo error rate are negatively associated with the Go hits response time (**Figure 4.3**). The longer the response time is, the less the commission errors are. The significant correlation between those two measures suggests that quicker responses may be associated with certain form of diminished executive control (more discussion will be provided in the following section).

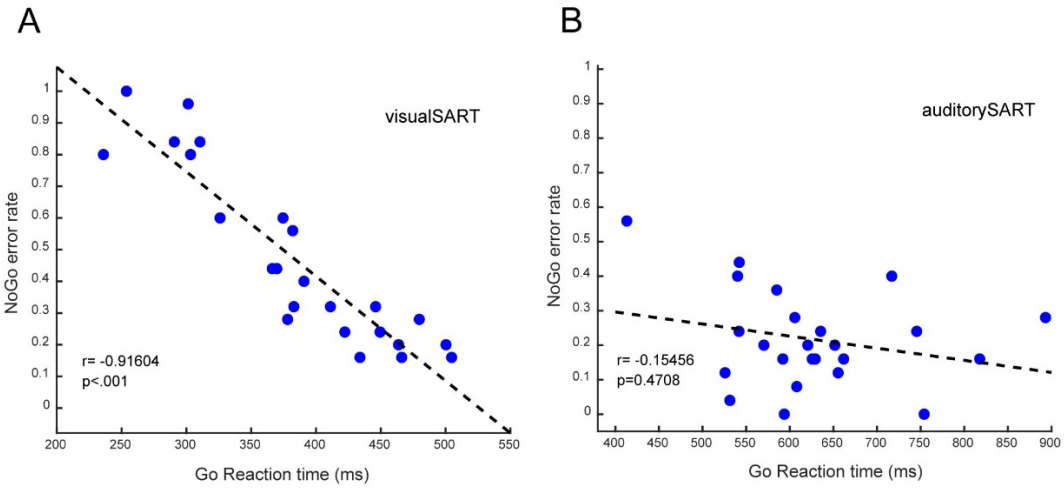

Figure 4.3. The spearman correlation analysis was applied on both tasks, all individuals were included (N=24). (A) Significant correlation between Nogo commission error rate and Go hits response time in visual SART was observed. (B) No significant correlation between Nogo commission error rate and Go hits response time in auditory SART.

The coefficient of variation (CV) of response times (RTs) is hypothesised to reflect fluctuations in RTs, indicating episodes of speeding and slowing due to lapses in attention (Seli et al., 2012b). The CV of RTs were analysed in separate groups: a full group of 24 participants [t(23) = 8.389, p < .001] of and a selected group of 19 participants who had a commission error rate below 80% in the visual SART [t(18) = 8.171, p < .001]. In both groups, significant differences were observed in the CV of RTs between modalities, indicating task-dependent variability linked to attentional lapses (**Figure 4.4**).

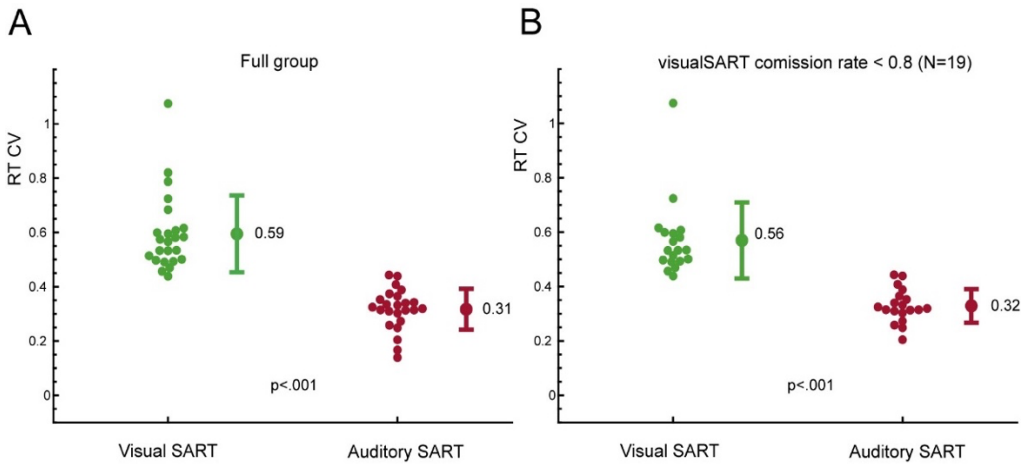

Figure 4.4. Coefficient of variation of response time (RTCV). (A) Distribution of individual performance and mean of RTCV across all GO trials for each of auditory, visual SART (p<.001). All subjects were included (N=24). (B) Performance of coefficient variation of RT across all GO trials for subject of those visual SART commission error rate below 80% (N=19) (p<.001).

Spearman correlation was run to assess the relationship between CV computed during visual SART and auditory SART. The analysis suggests that there was a statistically significant, positive correlation between those two measures [r(19) =.57,p=.01] (**Figure 4.5**).

Except for assessing the consistency across modalities. A significant correlation was seen between RTCV and commission errors in visual SART (selected group, N=19) [r(18)=0.5152, p=0.024]. The effects are stronger in full group (N=24) [r(23) =0.6853,p<.001]. However, correlation was not seen in auditory SART in both groups.

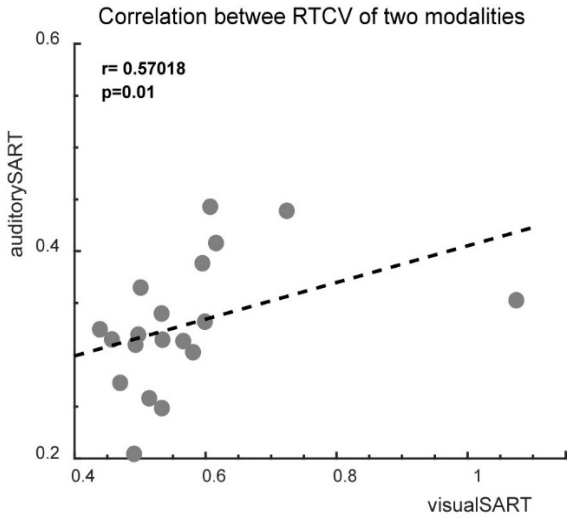

Figure 4.5. Spearman correlation. Significant correlation (r(18)=0.57,p=.011) was seen in RTCV between auditory SART and visual SART after the rejection of the subjects whose commission error rate is higher than 80% in visual SART (N=19).

4.5. Discussion

This sub-study closely replicates the observation from the previous study (Seli et al., 2012a). Firstly, significantly improved performance with less commission error rate was seen in auditory SART, compared with visual SART; meanwhile, distinctly longer response time appear in auditory version. In order to assess whether the slower RTs account for the reduced commission errors/better performance, the study revealed a significantly negative correlation in visual SART. Additionally, correlation between the two task versions in coefficient of variance of RTs (RTCV) was only observed in selected group, which is different from the results of previous study that they tested on full group. This discrepancy is likely to be attributed by potential noise from the unsupervised online environment, leading the analysis to exclude outlier performances in the visual SART.

4.5.1. What Contributed to the Prolonged Response Times Observed in Auditory SART Go Trials

Unlike visual stimuli, which have a consistent duration of 250 ms, auditory stimuli exhibited some unavoidable variations in temporal duration, averaging 0.35 ± 0.0516 s. This longer duration of most auditory stimuli compared to visual stimuli could explain the extended Go trial response times seen in auditory SART. When assessing the mean response time differences between the two versions of SART, it was found that the auditory SART was about 240 ms slower, suggesting that the duration differences between the stimuli might not be the only predictor of response time differences (Cheyne et al., 2009). Auditory stimuli, unlike visual stimuli (which are presented holistically), require a longer processing time due to their sequential nature. This provides a wider time window for the responses, potentially allowing participants more time to recover from transient attention lapses (Weissman et al., 2006). This aspect, alongside the speed trade-off discussed later, supports the hypothesis that participants were more optimally prepared with auditory tasks, as reflected by the prolonged response times.

4.5.2. What does the Commission Error Reflect

Scientists regard the underlying cause of commission errors (fail to withhold the response to Nogo trials) as sustained attention lapses. In the mindlessness theory, lapses of sustained attention are attributed to the subjects' withdrawal of their conscious attention

from monotonous task and redirect their attention to task-irrelevant thoughts or mind wondering (Giambra, 1995; Smallwood et al., 2004). Instead, in the resource theory, the perspective argued that lapses of sustained attention are primarily attributed to subjects' mental fatigue and the depletion of their limited attention resources (Helton et al., 1999, 2005, 2009; Temple et al., 2000; Warm et al., 2008). Nevertheless, all the debate are built based on the reliance that the causes of SART commission errors is sustained attention lapses.

Helton (2009) argued that the SART might tap into processes other than just sustained attention, such as impulsivity and response strategies. The repetitive nature of responses to common signals in SART might encourage a ballistic feed-forward motor program that the supervisory attention system struggles to control, particularly when multitasking or distracted. For example, the team observed that participants were consciously aware of the incorrect response they made during the SART, but cannot prevent themselves from pressing the button when the speed was too fast. Thus, it was suggested that actively slow response can inhibit this self-assembling feedforward motor program. In other words, the hypothesis characterised that the commission error of SART is measuring the control ability of this supervisory system. Although one plausible explanation is that this constant regulation of the motor response can be regarded as one form of sustained attention (Helton, 2009).

4.5.3. Why Reduced Commission Errors were Performed in Auditory SART

This sub-study replicated the previous finding that the performance of auditory SART is better than visual SART regarding to the commission error (Seli et al., 2012a). Providing the context, our results suggested that this improved accuracy (reduced commission errors) are linked to the reduced Go trial response times (Figure 4.3). Critically, the significance is only observed in visual SART. The overall longer response time in auditory SART cannot explain the variability of commission errors. This might be attributed by the fact that SART is sensitive to the response strategy and subject to the speed-accuracy trade-off (Temple et al., 2000; Peebles and Bothell, 2004; Helton et al., 2009, 2010).

For example, one previous study addressed that a simple alteration in instructions emphasizing a slow-and-accurate rather than a fast-and-accurate strategy, producing substantial improvements in SART performance (Dang et al., 2018). Additionally, evidence was found to suggest that the commission errors measured in SART reduces with increasing age, but this reduction is accounted for by the robust slowing of RTs with increasing age (Carriere et al., 2010). This is consistent with the earlier theory model, which suggested that the responding process of SART should be characterised by three segments. The first

segment represents the participants simply reacting to the onset of the stimuli; the second segment are the ones characterised by longer response time and improved accuracy (stimuli is fully aware and maintained perceptually); the third segment represents the longest response time, during which the participants were adapting strategy for achieving high accuracy (Wood and Jennings, 1976).

One important question is whether the dynamics of the speed-accuracy trade-off are neurally linked to sustained attention. Sustained attention is not a singular cognitive function but a multifaceted process involving multiple sub-components and neural systems. Key regions such as the prefrontal cortex, anterior cingulate cortex, and parietal cortex play critical roles in this process (Fortenbaugh et al., 2017; Esterman and Rothlein, 2019). The speed-accuracy trade-off reflects the cognitive and motor processes of balancing quick responses with accurate performance. Variations in sustained attention could potentially interact with corresponding neural networks, thus impacting the dynamics of the speed-accuracy trade-off. Given the above, it is crucial to consider the variability in neural processes associated with the speed-accuracy trade-off when examining sustained attention.

4.5.4. Visual SART or Auditory SART

Response time coefficient of variation (RTCV) is a sensitive variable that measures the fluctuations in response time, thereby indicating the episodic lapses of attention. This sub-study found higher RTCV in visual SART (mean = 0.59) relative to auditory SART (mean = 0.31). This discrepancy may be attributed to the faster processing duration of the visual stimuli, which makes it more sensitive to moment-to-moment attention lapses (Cheyne et al., 2009).

Furthermore, the commission error rate was positively correlated with RTCV in visual SART. This observation can explain the large variation in response times: when experiencing attention lapses, the participants respond rapidly like an automatic robot, leading to errors. After an error occurs, error feedback results in post-error slowing of responses. The relatively slower stimulus processing duration and response rate in auditory SART makes the task more tolerant to these effects, hence smaller RTCV and commission error rates.

One of the purposes of using SART is to investigate whether the attention lapses account for the variability observed in pattern detection performance. It was found that the inherent properties of the auditory modality made it easier for participants to achieve high accuracy, thus limiting its ability to detect individual variability in attention lapses of healthy listeners. Therefore, a rapid response rate might be a more sensitive measure of minor

fluctuations in sustained attention, which makes the visual SART a better assessment tool for attention lapses and for revealing variability in explicit pattern detection performance.

4.6. Sub-study 2: Cognitive Underpinnings of Auditory Pattern Detection

4.7. Introduction

In the behavioural experiment of Study 2 (Chapter 3), participants were tasked with identifying the emergence of a regular pattern from random sound sequences, while silent gaps of 100 ms, 200 ms, and 500 ms were introduced between tone-pips. These gaps resulted in pattern durations of 1500 ms, 2500 ms, and 5500 ms, respectively. When the pattern duration was extended to 2500 ms (Gap200), participants still maintained high performance with a mean d' score above 2. However, as the gap was up to 500 ms (Gap500), resulting in a pattern duration of 5500 ms, overall performance dropped significantly, with a mean d' score of 0.97. The results also revealed considerable variability: some participants achieved near-ceiling performance, demonstrating a strong capability to explicitly detect the slow pattern despite its challenging nature. In contrast, nearly half of the participants performed at chance level, indicating their inability to identify the pattern.

Monitoring slow auditory patterns is indeed a challenging task. It demands substantial cognitive effort and energy due to the increased processing needs over extended timespans. This complexity also stems from the need to handle and manipulate auditory information over prolonged periods, which calls for various neural systems. Grasping which cognitive elements or computational procedures account for this variability is key to illuminate the neural mechanisms at play when listening under challenging conditions. Despite its obvious significance, there are still unanswered questions around which specific cognitive factors influence explicit auditory pattern detection and how they tie in with behaviour performance.

In this study, it is hypothesised that several cognitive factors may explain the variability observed in the pattern detection task. Specifically, both the Gap100 task (fast, tone presentation rate = 6.7 Hz) and the Gap500 task (slow, tone presentation rate = 1.8 Hz) were included to investigate whether cognitive abilities remain consistent across different time scales for auditory pattern detection tasks.

4.7.1. Auditory Working/short-term Memory

Auditory working memory refers to the ability to hold and manipulate auditory information temporarily during a cognitive task. Its key functions include: 1. Actively maintaining and manipulating auditory information. 2. Using executive processes to manage this information. 3. Linking auditory information to behaviour (for example, solving a task-related problem or following task instructions). These functions can be grouped into two main components. The first is the central execution, which directs attention and controls cognitive processes. The second is the phonological loop, which temporarily stores verbal/auditory information. (Baddeley and Hitch, 1974; Daneman and Carpenter, 1980). While both of those two components are essential, some studies suggest that the phonological component of working memory, rather than executive functions, plays a predominant role in various cognitive tasks requiring memory functions, such as the speech-in-noise (SiN) task (Millman and Mattys, 2017; Lad et al., 2020). Nevertheless, Bianco and Chait, (2023) presented compelling evidence that auditory working memory ability is not correlated with SiN task performance (Bianco and Chait, 2023).

The phonological component of working memory appears to be closely linked to auditory short-term memory, which primarily relies on activation patterns in the temporal lobe and less on executive functions (Cowan, 2019). Unlike working memory, the short-term memory is traditionally defined as the ability to temporarily retain auditory information and constrained by low-level sensory representation (Cowan, 2008). Nevertheless, several established models (Atkinson & Shiffrin, 1968; Cowan, 2008; Harrison et al., 2020) propose that this memory process also comprises multiple stages. Initially, information is automatically encoded and stored in a high-fidelity auditory memory buffer that preserves sensory details. This information is then transferred to a short-term store involving more active cognitive processes. Factors such as deliberate encoding strategies (e.g., elaborative rehearsal, chunking), domain expertise (e.g., musical expertise), and attentional resources can influence the efficiency and precision of this encoding process (Talamini et al., 2017). Although the shared neural mechanisms underlying auditory short-term memory and working memory remain debated and may evolve with new findings, this study does not rigorously distinguish between the two concepts. Instead, I will refer to the memory component involved in this task as short-term memory and will focus on the essential cognitive processes relevant to the task.

The auditory short-term memory is believed to be crucial in auditory scene analysis, functioning as a temporary repository for integrating and retaining sensory information (Sussman, 2005). Auditory streams often consist of multiple elements arranged in specific sequences, such as phonemes in speech or notes in music. The short-term memory allows the integration of these elements into cohesive patterns, aiding in the computation and

interpretation of auditory sources or meanings (Winkler and Cowan, 2005; Cowan, 2008). Temporal processing is another critical aspect, as timing and rhythm are essential cues for comprehending auditory scenes (Shamma, 2001). The short-term memory functions in perceiving temporal relationships between sounds, enabling individuals to discern beats in music or parse speech with proper cadence and intonation (Gussenhoven, 2004; Mauk and Buonomano, 2004; Hasson et al., 2015).

This sub-study employed a modified version of the classical auditory short-term memory test, 'Tone Pattern Comparison task' (TP-COMP; **Figure 4.6**) (Schulze et al., 2011; Albouy et al., 2013; Graves et al., 2019; Bianco and Chait, 2023) to investigate whether this task performance share variability with explicit pattern detection performance or not. Participants were instructed to memorize a 500 ms tone pattern (comprising 10 random 50 ms tones), retained it for 2 seconds, and then compared it to a subsequent probe sound pattern. Although structurally similar to the digit span task (Richardson, 2007; Woods et al., 2011), a traditional measure of active auditory short-term memory, TP-COMP uses rapid, arbitrary tone patterns that prevent rehearsal and influence of long-term memory, allowing us to measure low-level short-term sensory representations (Bianco and Chait, 2023).

4.7.2. Sustained Attention

It is hypothesised that maintaining attention over prolonged periods is critical for identifying patterns, especially in slow sequences like those featured in our Gap500 task.

There are several reasons why sustained attention might play a significant role in auditory pattern detection. First, attention functions as enhancing the precision of sensory representation, facilitating the encoding of auditory signals during information processing (Pessoa et al., 2003; Odegaard et al., 2016; Mehrpour et al., 2020). Sustained attention, by engaging the brain's executive control network, allows listeners to constantly allocate more computational resources to encode or manipulate the target stream while ignoring or inhibiting irrelevant sounds, such as background noise in the room (Pessoa et al., 2003). Furthermore, evidence suggests that stream formation in auditory scene analysis is also influenced by attention. Shamma and colleagues proposed that attention can enhance feature salience, modify neural representation, and increase phase coherence among neural populations, thereby sharpening the perceptual boundary between the attended stream and the background (Shamma et al., 2011).

Therefore, in this sub-study, the classical visual SART (Sustained Attention to Response Task) was implemented, as introduced in sub-study 1, to investigate whether it explains variability of explicit pattern detection performance.

4.7.3. General Task Engagement and Vigilance

Task engagement is a critical factor influencing performance. One hypothesis posits that variability in the pattern detection task is closely linked to participants' level of motivation, given that those experiments were conducted online. To quantify participants' engagement and motivation, the Frequency Sensitivity Test (FST) and response times to STEP changes (STEP RT) were utilised. Response times to STEP changes serve as an extra metric for gauging attentiveness and motivation. Although fundamental processing speed is a factor, quicker response times may also indicate increased motivation. This is because motivated individuals typically apply their mental effort optimally to maintain high vigilance, leading to faster detection of stimuli. Likewise, the Frequency Sensitivity Test (FST) can also measure motivation. This test asks participants to differentiate between tones with slight frequency differences, a task that is designed to be simple for healthy listeners. Motivated individuals often put more effort into staying focused and adhering to instructions accurately, while those less motivated may exhibit decreased consistency and accuracy due to distractions.

Beyond assessing motivational impact, the FST ensures that online participants possess normal auditory function and can accurately discriminate pitch variations. This is crucial since frequency discrimination correlates with phonological processing, which is fundamental for recognizing sound patterns such as spoken language (McArthur and Bishop, 2004; Hill et al., 2005). Impaired frequency discrimination is commonly associated with dyslexia (Baldeweg et al., 1999; Banai and Ahissar, 2004).

4.8. Methods

4.8.1. Participants

224 participants recruited through prolific (www.prolific.co) took part in the study. Data from 109 subjects are included in the analysis below (35 females; average age, 24.5 ± 4.69 years). Data from 3 subjects were rejected due to failure to respond to STEP trials or because responses to STEP trials were too slow (same rejection criterion #3 as Study 2). Data from 47 did not proceed to the main task due to not passing the pre-determined performance threshold in the practice task (same rejection criterion #2 as Study 2). Additionally, about 29% of the participants who initially accessed the experiment did not pass the headphone screen and therefore did not proceed further (Milne et al., 2020). This fail rate is similar to the previous two studies and those reported in Milne et al. (2020).

4.8.2. Pattern Detection Task

Participants completed two types of pattern detection task (those were named as Gap100 and Gap500 below). The task and stimuli are same as those used in behaviour experiment of Study 2. In all cases tone pips were 50 ms long with frequencies drawn from a fixed pool of values with twenty frequencies (logarithmically spaced values between 222 and 2,000 Hz; 12% steps). The order in which these frequencies were successively distributed defined different conditions, that were otherwise identical in their spectral and timing profiles. RND sequences (50% of trials) consisted of randomly ordered tone pips. RNDREG sequences (50% of trials) contained a transition from a random (RND) to regularly (REG) repeating patter. REG sequences were created by drawing 10 different, randomly ordered tone pips and repeating that order 3 times. Novel stimuli were generated for each trial. To stay consistent with Study 2, the target trial (RNDREG) used in this study consisted of 60 tones with the transition from RND to REG always at the 30^{th} tone-pips. The primary purpose of keeping the transition time fixed is to ensure all trials contain the same amount of information before the transition. It is acknowledged that the potential confound of temporal expectation; however, it is believed to be unlikely a major issue, given the limited number of trials that experienced by participants. Two control trial types were also included: sequences of tone-pips of a fixed frequency (CONT) that lasted 4000 ms, and sequences with a step change in frequency partway through the trial (STEP: the change always occurred after 2000 ms). The STEP trial was for same purpose as previously explained, no silence gaps were added in CONT and STEP in this study as well. Importantly, as STEP/CONT stimuli were embedded within the pattern detection blocks and served as a good proxy for subject vigilance/engagement. It was therefore included among the “cognitive factors” for the following analysis. All sounds were generated in MATLAB with a sampling rate of 44.1 kHz.

Each pattern detection task included 40 trials: 15 RNDREG, 15 RND, 5 CONT, and 5 STEP. In the Gap100 task, 100 ms silent gaps were added between tone-pips in RND and RNDREG sequences, each lasting 9000 ms. In the Gap500 task, 500 ms silent gaps were added, extending the sequence to 33000 ms. Participants were asked to respond as quickly as possible by pressing the space bar when they detected a RNDREG transition or STEP change. Feedback regarding response time, expressed as the number of elapsed tones between the transition and the participant's button press, was provided at the end of each trial. Performance was measured using response time and d Prime(d’).

Same as in the behaviour experiment addressed in Study 2, the two tasks were presented in a fixed order to ensure that participants could practice sufficiently with the easier task (100 ms gap) before moving to the more difficult condition. This approach was intended to minimize the impact of unfamiliarity with the basic task rules, ensuring that any variability observed in the challenging condition would not arise from a lack of familiarity with

the tasks themselves. Following the two pattern detection tasks, participants proceeded to complete three cognitive tasks (in random order).

4.8.3. Visual SART

The same version of the visual Sustained Attention to Response Task (SART) as in sub-study 1 was employed. Digits were displayed at the centre of a computer screen in one of five randomly assigned font sizes—48, 72, 94, 100, and 120 points—corresponding to digit heights ranging from 12 to 29 mm. The task involved presenting 225 individual digits, each shown for 250 milliseconds, followed by a 900-millisecond mask. This mask consisted of a 20 mm ring with a diagonal cross at the centre. The presentation followed a paced onset-to-onset interval of 1150 milliseconds. Both the digits and the mask were set in white against a black background. The primary outcome measured was the percentage of failures on no-go trials (commission errors).

4.8.4. Frequency Sensitivity Test (FST)

Participants in this task must determine whether pairs of tones, separated by a 500 ms gap of silence, are the 'same' (50% of trials) or 'different' (50% of trials). The stimuli comprise 50-ms tone-pips, gated on and off with 5-ms raised cosine ramps. The frequencies are drawn from the same pool as the pattern detection tasks, using only contiguous pairs of frequencies in each 'different' trial. Each participant heard 39 pairs of tones in total. Participants were instructed to respond by pressing button 'S' for 'same', and button 'D' for different after heard each tone pair. The outcome measure for this task is the rate of correct responses. Despite the task's simplicity, it demands normal hearing and engagement to the tone pairs. The performance achieved by listeners is taken as an indicator of healthy listeners and also the general task engagement. (Bianco and Chait, 2023)

4.8.5. Tone Pattern Comparison Test (TP-COMP)

Stimuli. Each trial lasted 3000 ms and included two sequences of tone-pips, each lasting 500 ms and containing 10 tones. The pair of sequences was separated by a 2000 ms silent gap (see **Figure 4.6**). Each tone-pip was 50 ms long, and their frequencies were randomly sampled from the same pool used in the pattern detection tasks for consistency. The sequence before the gap represented the memory array, while the sequence after the gap was the test array.

Task. Participants were required to decide whether two sequences were identical or different by pressing button 'S' for "same" and 'D' for "different". The task comprised 32 trials, where each trial presented a memory array (encoding phase) followed by the 2 second silent gap (maintaining phase), and then the test array (retrieval phase); these arrays were identical on half of the trials and differed on the other half through the shuffling of three tones' positions. To prevent primacy and recency effects, the positions of first and last items were always unchanged. Novel stimuli were generated for each trial using MATLAB at a 44.1 kHz sampling rate, ensuring that each sequence was unique to avoid familiarity effects. The outcome measure was the correct response rate, assessing participants' ability to accurately discriminate two sequences (Bianco and Chait, 2023).

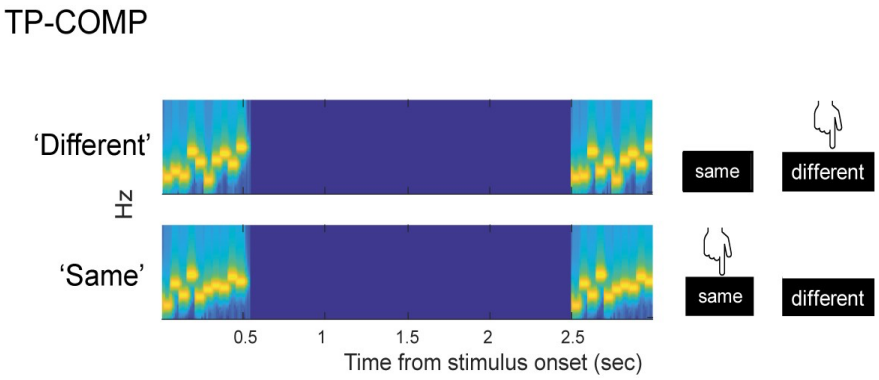

Figure 4.6. The spectrogram of the stimuli in the tone pattern comparison task (TP-COMP). Sequences of each trial consists of ten 50 ms tone-pips separated by 2 sec silent gap. Yellow square represents each tone-pip, purple area represents the silent gap. For those 'Different' trials, three tones' positions were altered, subjects were instructed to press the button once they have made the decision.

4.9. Results

Figure 4.7A displays distribution of d' performance on the pattern detection tasks from the full group of participants (N=109). Repeated measure ANOVA tested on d' indicates a significant difference between task conditions Gap100 and Gap500 [$F(1, 108) = 86.952$, $\eta 2 = .446$, $p < .001$], which replicated the results that have observed in behaviour experiments of Study 2.

Pattern Detection Simulation (d' modelling):

It is important to note that due to the extended duration required to complete the Gap500 task, the test block was limited to include only 15 target trials (RNDREG) and 15 foil trials (RND). This restricted number of trials inherently results in a wider distribution of chance performance. To assess the impact of trial quantity on performance, the d' scores of 109 subjects simulating random performance across two scenarios were modelled: one with 15 trials and another with 100 trials per condition (target or foil). The model assumes that participants respond to the trials randomly, irrespective of the number of target trials in the test block. The d' score is calculated based on each simulated participant's random response.

The d' score distribution is displayed in **Figure 4.8**, as expected, variability in d' reduced significantly with an increased number of trials. This implies that participants who cannot discern the pattern in the task might achieve d' scores below or above chance. Thus, performance above chance level (d' = 0) is intermixed with noise, and a d' score of 0 cannot serve as an effective threshold for distinguishing performance. While adding more trials could reduce noise, it would also result in participant fatigue due to the longer task duration. This limitation of behavioural measures, at least for d', poses inherent constraints. **Figure 4.8**C displays a comparison between human performance on the Gap500 task and the model simulation. A one-sample t-test indicates significant differences between these two distributions [$t(1,108)=10.245$, $d=1.16$, $p<0.001$]. This implies that the human participants were not simply performing at random. For the subsequent analysis, participants who achieved d' scores above 0.5 were selected (The 90th percentile was used as a cut-off point in terms of the random behaviour distribution from model simulation). 76% of the full group (N=109) passed the threshold and was advanced to the subsequent analysis.

For response time analysis, the study only focused on those participants who achieved a sufficiently high accuracy (d' > 2) so that their response time is interpretable (**Figure 4.7**B). Similar to the response time correction in Study 1 (Chapter 2), raw response times were corrected by subtracting the RT to STEP change and then converted to express the RT in terms of number of tones. Since the effective transition is theoretically detectable after the beginning of the second cycle, it is expected that the $RT_{number\ of\ tones}$ to be more than 10 tones (one cycle of pattern). Our results demonstrate that participants needed an average of 18.5 tones after the transition to detect REG in the Gap100 task and about 19.6 tones in the Gap500 task. Independent-t test confirmed that the response time was significantly slower in the Gap500 task relative to the Gap100 task [$t(78) = -2.7208$, $p = .008$], suggesting that participants required more information to determine the emergence of a pattern in the Gap500 condition.

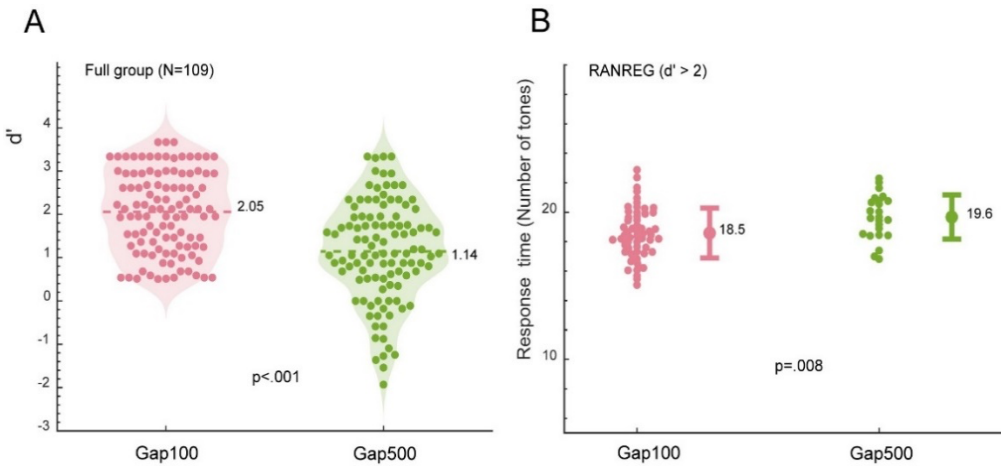

Figure 4.7. Performance(d') and RT $_{\text{number of tones}}$ to pattern detection tasks for all participants (N = 109). (A) Violin distribution of d' in the Gap100 task and Gap500 task. d' in Gap100 task is significantly higher than d' in Gap500 task, and increased individual variability was seen when silence gap is increased to 500 ms. The dashed line represents the mean performance. (B) $RT_{\text{number of tones}}$ of pattern detection tasks (d' > 2).

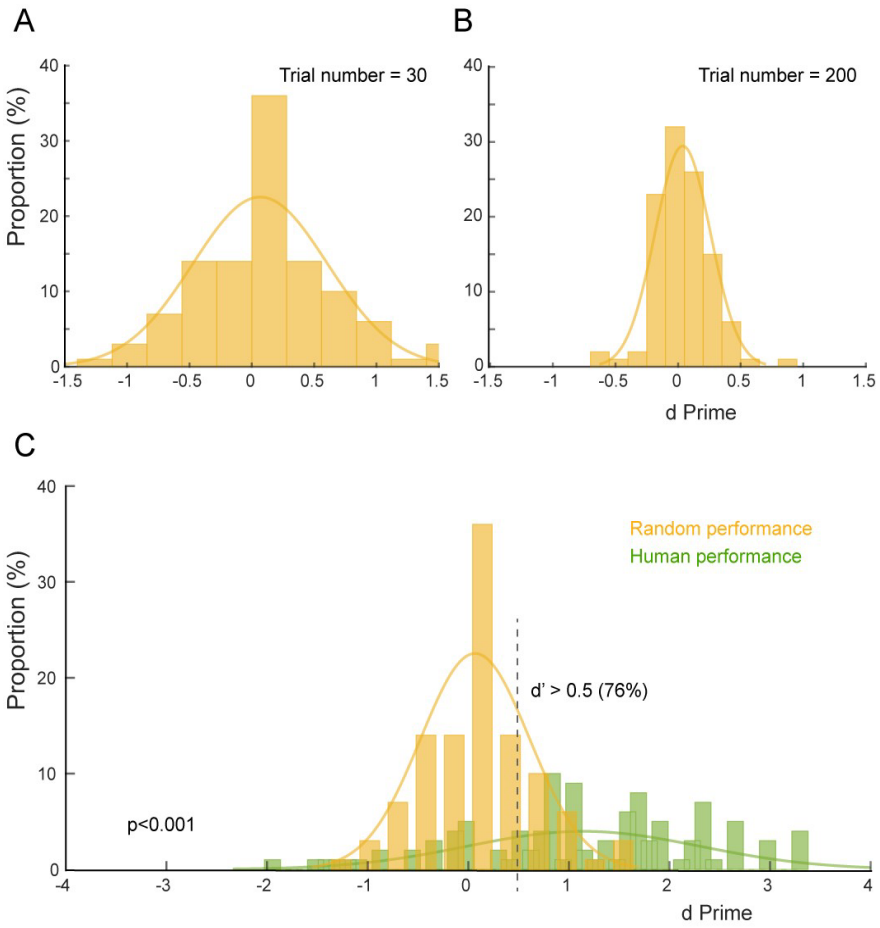

Figure 4.8. The simulation of pattern detection task performance modelled on the size 109 subjects. The distribution of d' was plotted for each trial number condition. (A) This plot shows the simulation for a trial number of 15 for each condition. The performance distribution is normally distributed, with d' ranging from -1.5 to 1.5, and the mean and median centred near 0. (B) This plot shows the simulation for a trial number of 100. Like the trial number of 15, the performance is normally distributed with the mean and median centred at 0. However, d' ranges from near -0.5 to 0.5, a significantly narrower range than in the trial number of 15. (C) d' distribution of Gap500 tasks from human subjects compared to the model simulation, significant difference was seen between two distributions ($p<0.001$). A threshold of 0.5 was chosen to screen out participants who might perform at chance. Approximately 76% of participants fall above this threshold.

To quantify cognitive task performance, the commission error rate on Nogo trials for the visual SART and the correct response rate for both the TP-COMP and FST tasks (**Figure 4.9**) were measured. Unsurprisingly, the data revealed substantial variability across all cognitive measures. But can any of this variability explain individual differences in pattern detection performance? To address this question, the study employed the multiple linear regression model with d' scores in the pattern detection task as response variable, performance on the TP-COMP, FST, SART tasks, and STEP change response time as the predictors, to investigate whether/which cognitive factor could explain the observed variability in d'.

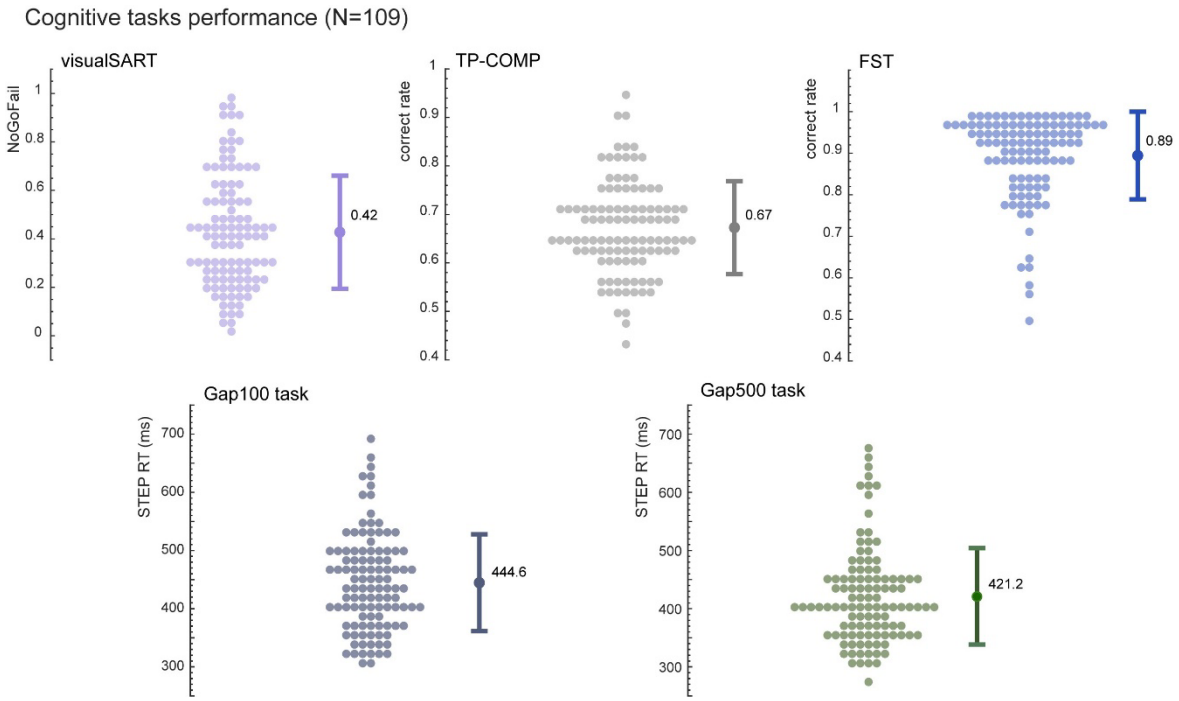

Figure 4.9. Substantial variability was observed from all cognitive measurements. The distribution of the outcome measures from tasks TP-COMP, FST, visual SART, and STEP RT from Gap100 task and Gap500 task. Each dot represents the individual performance.

The full group (N=109) in the Gap100 task was analysed since all participants performed above chance and d' is therefore interpretable. An additional point of interest is that this condition is quite fast, allowing to assume a degree of automaticity in performance

(i.e people have reduced ability to explicitly track the pattern). The model was tested on the d'Gap100 and cognitive measures (visual SART, STEP RT, TO-COMP, FST). The output revealed that the independent measures significantly accounted for 15% of the variability [R^2=0.15, F(4,104)=4.597, p=0.002]. The regression expression is as follows: d'Gap100 = −0.987 + (0.02*STEPRT) - (0.211*SART) + (0.208*FST) + (0.254*TP-COMP). Specifically, TP-COMP (p = 0.007), visual SART (p = .022) and FST (p = .025) contributed significantly to the model.

A similar analysis was conducted on the performance in the Gap500 task, focusing on the subjects who achieved above chance performance (d' > 0.5).

The model with d'_{Gap500} as the dependent variable, performance on the TP-COMP, SART, and FST tasks along with the STEP response time in Gap500 test block as predictors demonstrated that the independent measures explained 12.8% of the d' variability of d'_{Gap500} [R^2 =0.128, F(4,77)=2.831, p=.03]. The regression equation was: d'_{Gap500} = -0.48 + (-0.103*STEPRT) + (2.97*TP-COMP) + (-0.055*SART) + (0.115*FST). Within this, only TP-COMP (p = .008) contributed to the model with significance while other measures did not. This suggests that the large variability seen in the Gap500 task is predominantly driven by variability in auditory short-term/working memory abilities of subjects.

It is hypothesised that participants who achieve ceiling performance in the Gap100 task are more likely to be highly engaged, familiar with the task, and exhibit high baseline performance. Thus, their performance in the Gap500 task is expected to predominantly reflect variability due to individual differences in tracking pattern across different time scales, rather than task familiarity or engagement. To minimize the irrelevant 'noise' and enhance the signal of interests in the d' distribution, the subsequent analysis focuses specifically on participants who demonstrated ceiling performance (d' ≥ 3) in the Gap100 task. (**Figure 4.10**). For each of these participants (N = 22), the difference between their d' performance on the Gap100 task and Gap500 task was computed. As suggested by **Figure 4.11**, there is a large variability in performance on the Gap 500 task. Some participants maintained a high level of performance whereas others exhibited near chance performance despite doing exceptionally well on the Gap100 task.

The next aim is to determine whether there is an interaction between task conditions (Gap100 and Gap500) and cognitive factors. Multiple regression model with $d'_{Gap500,\ Gap100\_Diff}$ as the dependent variable, performance on the TP-COMP, SART, and FST tasks along with the STEP response time in Gap100 and Gap500 test blocks as predictors demonstrated that the independent measures explained 69.6% of the d' variability of $d'_{Gap500,\ Gap100\_Diff}$ [R^2 =0.696, F(5, 16) = 7.314,p < .001]. The regression equation was: $d'_{Gap500\_Gap100\_Diff}$ = -9.673 + (-0.462*STEPRT_Gap100) + (-0.171*STEPRT_Gap500) + (0.615*TP-COMP) + (-0.159*SART) + (0.257*FST). Within this, TP-COMP (p < 0.001) contributed to most of the

variance with significance, confirming the critical role of short-term memory when the time scales of the stimulus were extended. Besides, the STEP RT in Gap100 task also contributes to the model with significance (p=0.009), revealing that smaller the differences (more consistency across tasks) between d'Gap500 and d'Gap100, the faster the STEPRT in Gap100 is (more engaging and motivated the participant was). However, this was not seen in STEP RT in Gap500 task, which might intriguingly stem from increased task complexity (e.g. more cognitive efforts and cognitive processes such as strategy applications were involved in the task) resulting in greater variability in STEP response times, therefore a less direct relationship between response times and the task performance can be statistically identified in that condition.

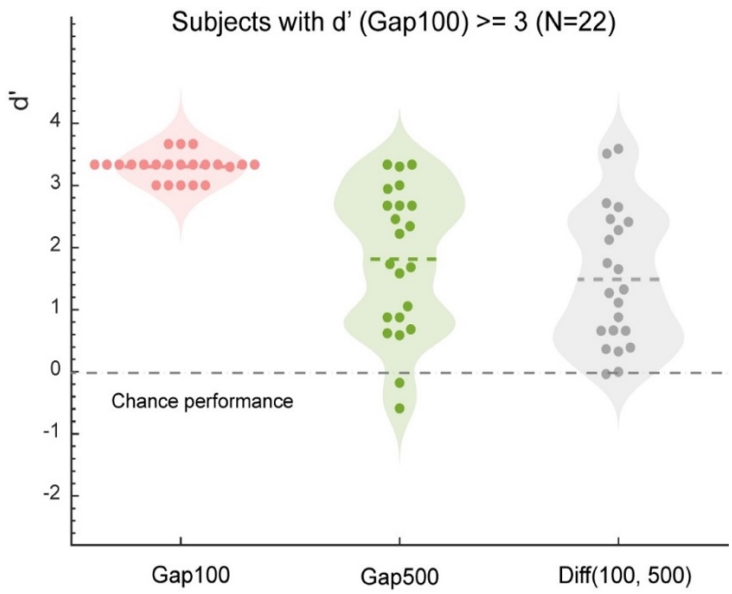

Figure 4.10. Large variability in Gap500 task were still observed despite the fact that all subjects can achieve ceiling peformance in Gap100 task. Violin distribution of d' in two tasks, and the d' differences. The coloured dashed line within the distribution represents the mean.

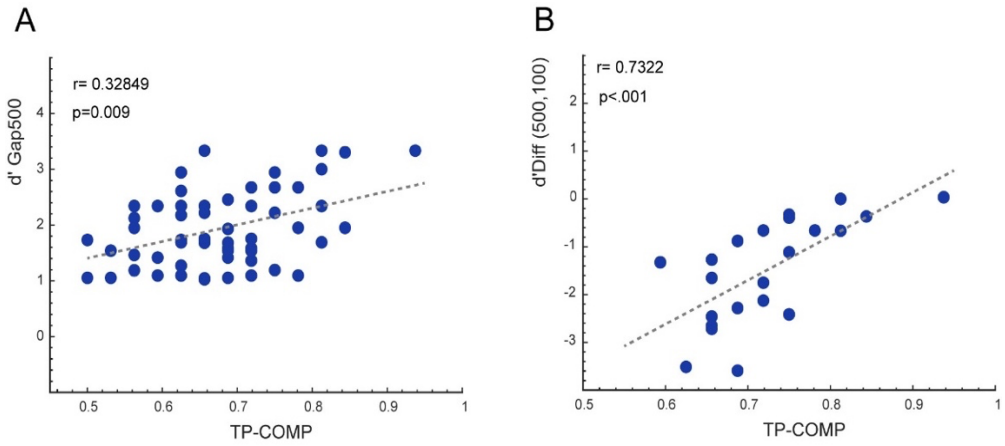

Figure 4.11. Spearman Correlation. (A) Participants with d' > 0.5 in Gap = 500 pattern detection (N = 82) were included. TP-COMP performance is positively correlated with d' in Gap = 500 ms ($p < .01$) task. (B) Participants with d' > 3 in Gap100 pattern detection task (N = 22) were included. The d' differences between Gap500 task and Gap100 task is positively correlated with TP-COMP performance ($p < .001$).

4.10. Discussion

This sub-study aimed to delineate the cognitive factors explaining the individual variability of pattern detection performance in the Gap100 and Gap500 tasks. The attention was particularly placed on the auditory short-term memory (TP-COMP task), the sustained attention (SART task), the basic auditory processing, as well as the general task motivation (FST and STEP RT). The findings provide novel insights into the differential contributions of these cognitive factors based on the temporal characteristics of the auditory pattern detection tasks.

4.10.1. Sustained Attention and Frequency Discrimination's Role in Gap100 Task Performance

Frequency Sensitivity Test (FST): The regression analysis showed that the Frequency sensitivity Task (FST) significantly accounted for individual variances in the performance of the Gap100 task ($p = 0.025$). This result is consistent with research by

Bianco and Chait (2023), who used a similar task to predict performance variance in sound pattern detection across various age groups. However, unlike their 20 Hz presentation rate, this experiment used a slower presentation rate (6.7 Hz, Gap100 task) of sound sequences. Despite these methodological differences, both studies suggest that the variability reflected by the FST (is hypothesised to be primarily reflecting the task engagement), which predicts the pattern detection performance across different rapid temporal scales (Bianco and Chait, 2023).

Critically, the impact of the FST was not significant in the Gap500 task performance. One possible explanation is that monitoring rapid sequences (Gap100) may require less cognitive effort compared to the slower sequences. The lower performance on the Gap100 task could be due to the reduced engagement, as this task—similar to the simpler FST task—does not demand extensive cognitive resources. However, the longer pattern task (Gap500) may exceed mere engagement requirements for successful performance. The substantial individual variability observed suggests that cognitive limits cannot merely be enhanced through increased motivation. Such interpretation is supported by the short-term memory, which is the solely significant predictor in the data. Moreover, given the extended duration of sound sequence in the Gap500 task, additional cognitive processes are likely engaged (i.e. strategy application). For instance, this task's increased mental demands may activate more complex neural strategies aimed at achieving specific goals. Such increased complexity introduces greater variability in performance outcomes, which in turn results in a less straightforward statistical correlation between the Frequency Sensitivity Test (FST) and d' scores in Gap500 task.

Sustained Attention to Response Task (SART): The SART also emerged as a negative predictor of Gap100 task performance ($p = 0.025$). The Gap100 task, being relatively fast paced, requires participants to monitor auditory stimuli continuously and respond swiftly. The significant contribution of sustained attention to performance aligns with previous findings (Helton, 2009), which demonstrated that the lapses in sustained attention impair performance in rapid auditory tasks. Moreover, Cheyne et al. (2009) emphasised that the visual SART is sensitive to moment-to-moment attentional lapses. Following this, it is possible that participants in Gap100 task may experience more attentional lapses due to the stimuli's fast-paced nature, thus impacting decision-making and execution processes (Cheyne et al., 2009). From the execution perspective, Helton (2009) also linked fast-paced tasks to the control ability of the supervisory system (executive network), particularly in situations where participants are aware of making incorrect responses but cannot inhibit them due to the task's fast speed (Helton, 2009). This inability to inhibit motor responses may explain some of the variability in Gap100 performance attributable to sustained attention. Alternatively, it is also possible that the shared variability is mainly accounted by the

heightened awareness that is required to monitor the rapid task such as Gap100. Since SART also demands constant vigilance to meet task objectives.

4.10.2. Auditory Short-Term Memory Predicts Performance in Explicit Sound Pattern Detection Across Two Time Scales

Auditory short-term memory, measured by the TP-COMP task, positively predicted performance in both the Gap100 and Gap500 tasks. This finding also aligns with Bianco and Chait (2023), which drew a similar conclusion. The result emphasised the consistent role of short-term memory across various temporal scales of auditory pattern detection, from a 20 Hz tone presentation rate (see Bianco and Chait, 2023) to a 1.8 tone presentation rate (Gap500 task).

Memory Processes in Gap100 Task: In the fast-paced Gap100 task, pattern detection requires listeners to maintain a sensory representation of the pattern that serves as evidence for comparative analysis. Given the presentation rate of the Gap100 sequence (6.7 Hz), which is relatively fast, it is hypothesised that detection relies largely on automatic tracking, where the pattern 'pops out' perceptually. The shared variability between TP-COMP and Gap100 performance suggests the involvement of a task-independent, low-level sensory memory component common to both cognitive processes.

Memory Processes in Gap500 Task: Unlike the Gap100 task, the presentation rate of the Gap500 task is slower (1.8 Hz). In this task, the TP-COMP significantly predicted the performance. It was the only cognitive factor among all measures that did so, highlighting the fundamental role of short-term memory in detecting slow patterns. Nevertheless, it is noteworthy noting the slower-paced nature of explicit pattern tracking, which suggested that multiple cognitive processes may contribute to the observed variability and interacted with the behavioural processes associated with TP-COMP.

For instance, cognitive strategies like auditory rehearsal has been found to be capable of enhancing task-related memory representations (Greene, 1987; Buchsbaum et al., 2005). Moreover, participants could deliberately track salient sounds in both the TP-COMP and Gap500 tasks rather than retaining entire sound sequences. This strategic tracking allows them to maintain pattern-relevant information over the longer intervals presented in the Gap500 task.

4.10.3. Implications for Neural Mechanism

The role of short-term memory in explaining individual variability in pattern detection task performance was consistently observed, as evidenced by the shared variability observed between the TP-COMP and both Gap100 and Gap500 tasks. This finding aligns with earlier research, such as that by Bianco and Chait (2023), which demonstrated correlations between the TP-COMP and tasks involving even more rapid pattern durations. Notably, the Gap500 task, which shows substantial individual variability, identified short-term memory as the only significant predictor. These results underscored the pivotal role of auditory sensory/short-term memory supporting the ability to detect complex patterns.

Beyond the short-term memory, it is likely that the observed relationship between the TP-COMP and pattern detection task could be explained by the involvement of high-level executive components of working memory, given that the TP-COMP task actively engages in the execution and manipulation of information (feature of working memory). This is particularly relevant for Gap500 task, given the auditory system's limited capacity to store sensory details for extended periods, resulting in an alternative approach for tasks that require prolonged temporal engagement (Keller et al., 1995). As discussed before, listeners might selectively attend to certain salient sounds within the sound sequence in both tasks.

There are two primary reasons why such adaptations are theoretically plausible and crucial for interpreting our results. First, the slower pace of tasks like the Gap500 provides a more extended time window, allowing the brain to engage higher-level neural pathways, such as rehearsal and logical reasoning. This was supported by the research in both human and animal models that areas like the dorsolateral prefrontal cortex (also involved in working memory) are more actively engaged during tasks that allow for these higher-level processes due to the longer task duration (Miller et al., 1996; Curtis and D'Esposito, 2003). In contrast, tasks that involve rapid sequences, where the swift presentation constrains the time available for higher cognitive processes, rely more on automatic sensory processing (Scott et al., 2006; Griffiths and Hall, 2012). In addition, as discussed previously, rapid sensory inputs are evolutionarily linked to situations signalling danger or requiring immediate response, thereby prompting the brain to activate attentional and vigilance mechanisms (Jasmin et al., 2019). Moreover, the constraints of memory capacity are evident as sensory representations decay over time (Hardt et al., 2013). In tasks requiring monitoring slower stimulus, this decay makes sensory information particularly vulnerable, forcing listeners to develop alternative strategies to maintain and manipulate this information to achieve task goals.

4.11. Conclusion

In summary, this sub-study provides compelling evidence that different cognitive factors are differentially implicated in auditory pattern detection tasks with varying temporal demands. In scenarios where auditory patterns are presented at a rapid pace, the predominant factors that predict performance revolve around task engagement, general vigilance, and short-term memory ability. This is evidenced by significant correlation with all testing metrics such as the FST, SART and TP-COMP.

Conversely, as the demands on efforts increase with slower pattern, the key determinant of performance shifts towards auditory short-term memory. This transition underscores the primary role that memory plays in the tracking of extended sequences. It indicates that when the task requires the retention and manipulation of information over longer periods, memory capacity becomes the major factor that contribute to the variability of the performance metrics, overshadowing the influences of motivation and attention.

However, since the interpretations are constrained by the behavioural nature of this study. As discussed above, the shared variability between TP-COMP and pattern detection tasks may extend beyond short-term memory capacity to include processes such as explicit information manipulation (as involved in working memory) and the application of cognitive strategies. To further explore these dynamics, my next study aim to delineate specific cognitive processes by recording brain activity under passive listening. I employed EEG to explore the neural correlates of the unsupervised pattern detection process, independent of top-down attention. This study further assesses the relationship between the neural correlates of auditory patterns and individual variability of short-term memory, as evaluated through TP-COMP measurements.

5. Chapter 5: Sensitivity to Complex Sound Patterns is Correlated with Auditory Short Term Memory

5.1. Introduction

The brain does not merely record the world passively; rather, it actively constructs a sensory representation of the perceived environment. Accumulating evidence including Study 2 suggests that the brain automatically attunes to complex auditory patterns across various time scales, exhibiting heightened sustained responses to predictable sequences as opposed to unpredictable ones (Barascud et al., 2016; Southwell et al., 2017; Southwell and Chait, 2018; Hu et al., 2024). This sensitivity has been hypothesised to reflect the brain's ability to encode the predictability of sensory input. More recently, this neural correlation has been substantiated by evidence as 'precision' — a key concept which is central to predictive coding theory (Zhao et al., 2024). As reviewed in the general introduction, precision is defined as the degree of inferred predictability of sensory inputs, effectively representing the inverse variance of the top-down predictive distribution (Friston, 2010).

To recognise sound pattern, the brain must formulate predictions about forthcoming sounds and test these predictions against the actual sensory input received. This cognitive process requires the brain to maintain an elaborate record of previous sound sequences, using the stored information to both construct and refine a generative model continually. Research indicates that the dynamics of sustained responses, supported by the network of auditory cortex, inferior frontal cortex and hippocampus, may serve as manifestations of this generative model (Barascud et al., 2016; Zhao et al., 2024).

Theoretically, the success of the model is dependent on the auditory memory's ability to preserve the necessary information. As Barascud et al, (2016) suggests, sustained response amplitudes gradually decline with increasing sound pattern complexity. This suggests that while the brain uses predictability in regular patterns to forecast auditory events, more complex (complexity here is defined as a greater number of tones are included within the pattern) and longer patterns potentially place greater demands on memory resources. As sequences become longer and more complex, the sustained response amplitude decreases. This correlation might reflect a greater work load required for information processing and memory retention, therefore serving as an indicative of memory capability (Barascud et al., 2016). Building on this, Southwell et al. (2018) provided EEG evidence from participants who listened passively to sound sequences that were either random or regular, punctuated by occasional deviant tones. Similarly, their findings showed

a more pronounced sustained response to regular sequences over random ones. However, deviations within these regular sequences evoked significantly stronger and more distinct neural responses than those in random sequences. This suggests that regularity may reduce the cognitive load needed to analyse stimuli, potentially easing the burden on memory capacity. These observations further forge a link between sustained neural responses and the brain's limited memory capacity (Southwell and Chait, 2018).

Furthermore, observations from Study 2 demonstrate that lengthening the auditory pattern duration from 500 milliseconds to 2500 milliseconds significantly diminishes the sustained neural response to those patterns compared to shorter ones. This differential response indicates that longer sequence durations might intensively challenge the memory system, thus offering insights into the dynamics of auditory memory and its capacity (Hu et al., 2024). Age-related differences provide additional support for this relationship. The study conducted by Herrmann et al. (2022) explored how aging affects neural responses to auditory patterns in younger and older adults. The team found that while older adults demonstrated heightened responsiveness to the initial sounds, their sustained neural activity to regular auditory patterns was significantly reduced compared to younger adults (Herrmann et al., 2022). The authors proposed that this reduction in neural activity suggests an age-related decline in the auditory system's ability to process and maintain information over time, a degeneration that perhaps corresponds with the memory commonly associated with aging (Gazzaley et al., 2005).

Beyond research on regularity, neural evidence from memory studies provides compelling evidence that sustained neural activity is closely linked to memory processes. For example, Curtis and D'Esposito (2003) reviewed the role of sustained neural activity sourced from dorsolateral prefrontal cortex (DLPFC) in working memory (WM). The single-unit recordings from the monkeys suggested that during the retention intervals of delayed response tasks, persistent, elevated levels of neuronal firing are observed in the DLPFC. FMRI findings on humans further corroborate this, showing sustained activity in DLPFC under similar task conditions. Notably, the study emphasised that both the duration and intensity of this sustained activity, relative to the number of items retained in memory, are crucial indicators of the DLPFC's processing efficiency in maintaining phase (Curtis and D'Esposito, 2003). Additionally, Axmacher et al. (2007) explores the neural mechanisms within the medial temporal lobe (MTL) that has been hypothesised to support WM functions. The researchers utilised intracranial EEG (iEEG) and fMRI to observe WM-specific sustained neural activity, analysing the effects of maintaining single or multiple items (photographs of faces) in memory. The study discovered that maintaining an increasing number of items induced both a negative shift in the direct current (DC) potential and an increase in gamma-band activity within the MTL. This observation suggests that those neural activity are modulated by the WM load, in particular, the sustained activity varied depending on the

number of items retained. Those findings provide direct evidence that the sustained activity measured in MTL which contains the hippocampus and parahippocampus, is inversely correlated with memory load (Axmacher et al., 2007).

Intriguingly, Kumar et al. (2021) systematically investigated the neural underpinnings of auditory working memory (AWM) by analysing local field potentials (LFPs) across various brain regions involved in auditory processing. Employing electrocorticography (ECoG), the researchers focused on the oscillatory activity within the auditory cortex, inferior frontal cortex, and the hippocampus during tasks that required maintaining sounds in memory. The study observed distinctive patterns of sustained neural activity, specifically noting an enhancement in delta and theta oscillations and a suppression of beta and gamma frequencies during the memory maintenance phase. Even though this study only required participants to retain a single tone over a delay period, it still identified the same critical network. The network, involving the auditory cortex, inferior frontal cortex, and hippocampus, is associated with sustained responses observed in regularity detection. (Barascud et al., 2016; Hu et al., 2024). This shared neural network shed light on the possible connection between those two processes (Kumar et al., 2021b).

Taken together, the accumulated evidence provides an insightful perspective into the potential association between sustained neural responses of regularity detection and memory functions. These neural activities are crucial for understanding the neural mechanisms through which the brain analyse and represent sensory information. Despite these advances, evidence linking these specific neural dynamics directly to memory capabilities has yet to be established.

5.1.1. The Goal of the Study

Auditory short-term memory is believed to be crucial in auditory scene analysis, functioning as a temporary repository for integrating and retaining sensory information (Sussman, 2005) Importantly, this memory component is also known to be limited by low-level sensory representation and is prone to decay over time. To this end, this study employed the Tone Pattern Comparison task (TP-COMP; **Figure 5.2**A), the same task that was used in Study 3 (See **Chapter 4**), specifically designed to assess the active components of auditory short-term memory for sound patterns. In this task, participants are required to memorize a 500 ms tone pattern composed of ten 50 ms tones, retain this information for 2 seconds, and then compare it to a subsequent probe pattern. The TP-COMP, which involves rapid, arbitrary tone sequences, essentially measures the encoding, retention, and retrieval phases of memory processing. Therefore, it is hypothesised that the memory function variability as measured by TP-COMP might explain those individual variability observed in

sustained neural responses to sound patterns (Richardson, 2007; Woods et al., 2011; Recasens et al., 2015).

Study 2 investigated the brain responses to repeated, regular patterns and random sound sequences across two time scales: fast rate (20 Hz, 0 ms inter-tone interval, ITI) and slow rate (4 Hz, 200 ms ITI). Behavioural assessment at 4 Hz reveals high sensitivity to the pattern emergence across all participants. Although the slower pattern elicited a diminished sustained response compared to the faster pattern which suggests memory limitation, the relatively stronger response to the patterned versus random sequence suggests that the memory demands of the task were manageable for those participants. This indicated that the 2500 ms pattern duration may not have been sufficiently challenging to elicit varied responses that could demonstrate significant differences in memory capabilities among individuals. Additionally, since Study 2 used a 200 ms inter-tone interval (ITI), analysis focused primarily on the early stages of auditory processing, such as the P1/M50 and N1/M100 components (Hu et al., 2024). However, observations from previous research suggests that later stages of phase-locked responses are also indicative of information processing and statistical learning (Näätänen, 1990; Paavilainen, 2013; Maheu et al., 2019). For instance, the knowledge of high-order community structure embedded within image sequences arise around 500ms after stimuli onset and well predict behavioural performance within trial (Ren et al., 2022); Moreover, ERP such as P300, which is commonly identified after 300ms post-stimulus onset, is suggested to be sensitive to both global and local statistical context in auditory sequences (Kolossa et al., 2015).

This brings us to the question of how to design the paradigm for this study. As observed in Study 2 and Study 3, the behaviour findings related to explicit pattern detection exhibited a substantial variation at a longer pattern duration of 5500 ms (Gap500 task). Performance differed significantly among participants, with some performing exceptionally well and others at chance. Likely, the regression analysis in Study 3 suggests that these differences originate from individual variations in auditory memory, as indicated by the Tone Pattern Comparison task (TP-COMP), which significantly predicted performance. This finding supports the idea of a shared cognitive mechanism between explicit pattern detection and auditory short-term memory. However, it is still unclear about the underlying neural mechanism that predominantly contribute to the variability since multiple processes could be involved. As discussed in Study 3, except for the known procedures of sensory encoding, retention, and information manipulation (attention involved) in the tone-pattern comparison task. Other higher-level cognitive processes, such as strategic application (the listener might monitor the salient sound to identify the change in two tasks), may also explain the correlation.

Building on these insights, the current study extended the examination to implicit pattern detection, and aims to explore the interplay between auditory short-term memory

capabilities and the neural correlates of pattern sensitivity in the absence of voluntary attention. A silent gap of 500 ms between tone-pips (presentation rate of 1.8 Hz) in both regular pattern (predictable) and random (unpredictable) sequences were introduced, extending the pattern duration up to 5500 ms. Naïve participants listened to these sequences while watching a movie of their choice, with EEG recording their brain signals. Post-recording, participants completed the TP-COMP task.

This study pursues two primary goals: first, to explore how short-term memory contributes to tracking predictability in scenarios where attention is not actively employed; second, to assess whether the neural mechanisms underlying later time intervals (200-500ms relative to tone onset) of tone-evoked activity are indicative of regularity encoding.

5.2. Methods

5.2.1. Experiment

5.2.1.1. Stimuli

Stimuli were sequences of 50-ms tone-pips, each gated on and off with 5-ms raised cosine ramps (Figure 5.1A). Frequencies were selected from a pool of 20 values equally spaced on a logarithmic scale from 222 to 2000 Hz (12% steps). Two sequence types were created: REG sequences were generated by randomly selecting 10 frequencies from the pool without replacement. These were arranged in a specific (randomly determined) order to form a pattern, which was then repeated three times. New REG patterns were generated for each trial. RND sequences also consisted of 10 frequencies, similarly, selected anew for each trial and presented in random order. All sequences contained 30 tone-pips, presented with a 500 ms inter-tone silent gap (1.8 Hz rate; 5500ms REG cycle duration; 16.5 sec overall sequence duration). Stimulus delivery was in blocks. Each block consisted of 10 REG and 10 RND trials. A total of 5 blocks (100 trials) were delivered, with each condition presented 50 times in a randomised order.

5.2.1.2. Procedure

The experiment was implemented in Psychophysics Toolbox in MATLAB (Kleiner, Brainard, Pelli, & Ingling, 2007) and conducted in a sound-proof booth. EEG signals were recorded by Biosemi system (Biosemi Active Two AD-box ADC-17, Biosemi, Netherlands) with 64 Ag-AgCl electrodes at a 2048 Hz sampling rate and subsequently downsampled to 256 Hz. The recording was restarted for each block. Auditory stimuli were delivered binaurally via tube earphones (EARTONE 3A 10 Ω; Etymotic Research) inserted into the ear

canal. The loudness was adjusted to each participant's comfort. The experiment lasted a total of 40 minutes.

The experimental session began with an auditory functional localiser block, lasting about 3 minutes. This block featured a randomised sequence of 180-200 pure tones (1000 Hz frequency, 150 ms duration), each followed by a random interstimulus interval (ISI) between 700 and 1500 ms. This process served as a control to ensure a reasonable signal-to-noise ratio.

In the main experiment, participants passively listened to randomly presented stimuli with an inter-stimulus interval (ISI) of 3000-4500 ms, while watching a silent movie of their choice. Unaware of the auditory stimuli's nature, participants were encouraged to focus on the movie. The session was divided into five 8-minute blocks, with short breaks allowed between blocks while participants remained still.

5.2.1.3. Participants

Thirty-four naïve participants with normal pure-tone thresholds (≤ 20 dB) in standard audiometric frequencies (0.25–8 kHz) participated in the study. Data from two participants were discarded due to excessive noise, and data from another two were discarded due to a trigger malfunction resulting in data loss. Hence, data from 30 participants (19 female; average age, 24.7 ± 4.63) are reported below. All subjects were fluent in English, had normal or corrected-to-normal vision, and were reimbursed for their time. They had no history of hearing impairment or neurological disorders. The research ethics committee of University College London approved all experimental procedures described in this study, and written informed consent was obtained from each participant.

5.2.2. Data Analysis

5.2.2.1. Detrending

Before preprocessing, robust detrending was performed using a 10th order polynomial fit (de Cheveigné and Arzounian, 2018), with the fitted values subsequently subtracted from the original continuous signal. This was done to reduce common drifts in EEG recording, especially since the stimulus used in this study is particularly slow. This approach (as opposed to high pass filtering) helps preserve the slow dynamics (sustained response) that are phase-locked to the stimulus and are of interest in this study.

5.2.2.2. EEG Data Preprocessing

All pre-processing and time domain analyses were conducted using the fieldtrip toolbox (http://www.fieldtriptoolbox.org/, (Oostenveld et al., 2011). Low-pass filtering was

applied at 30 Hz (all filtering in this study was performed using a two-pass, Butterworth filter with zero phase shift). To analyse time domain data, the 10 most responsive channels for each subject were identified. This was done by combining tone responses collapsed from all conditions and identifying the N1 component (80-120 ms) of the onset response (Näätänen and Picton, 1987; Stufflebeam et al., 1998). For each subject, the 10 most strongly activated channels at the peak of auditory N1 (5 most positive, 5 most negative) were selected to best represent auditory activity for all subsequent time-domain analyses. This procedure served the dual purpose of enhancing the relevant response components and compensating for any channel misalignment between subjects.

5.2.2.3. Sequence Evoked Response

First analysis focused on responses to the sequence, with particular interests on low frequency activity as a potential marker of predictability tracking (Barascud et al., 2016; Southwell et al., 2017; Hu et al., 2024). High-pass filter was not used in this analysis. The data was segmented into epochs of 16.5 seconds, starting from 200ms prior to onset until the end of each trial. These epochs were then baselined to the pre-onset interval (200ms) and averaged for evoked response analysis.

To minimize low frequency drift artifacts and enhance signal noise ratio that is locked to the trial, denoising source separation (DSS) was applied as those was done in Study 2, following the method (de Cheveigné and Parra, 2014), three most significant components, which exhibited the highest reproducibility across trials, were identified and projected back into sensor space for each subject.

5.2.2.4. Tone Evoked Response

A secondary analysis was focused on neural responses to individual tones in REG versus RND sequences. To identify activity linked to each tone-evoked response, which might be obscured by slow neural dynamics, the raw data were high-pass filtered at 1.5 Hz. The filtered data were then segmented into individual tone epochs, spanning from 50 ms before to 500 ms after tone onset. Denoising Source Separation (DSS) was applied to the tone-evoked responses and the three most significant components were projected back to the sensor space for each participant. Responses from tones within each cycle were then averaged, producing three time series. This was done for each condition per subject. The time series were baselined based on the activity before the onset of the tone (50ms).

5.2.2.5. Statistical Analysis

The time domain data is summarised as root-mean square (RMS) across ten most responsive channels for each subject. As suggested by Study 2, RMS is a useful summary signal because it reflects the instantaneous power of the neural response, regardless of polarity that reflected by electrodes. For illustration, this study shows the group-response

(average of individual RMSs) and standard error across subjects. However, statistical analysis is always conducted across subjects. To assess differences between conditions (RND vs REG), RMS differences were calculated at each time point for each participant. A bootstrap resampling (Efron and Tibshirani, 1998) with 1000 iterations were then applied to the entire epoch. A significant difference was considered if the proportion of bootstrap iterations that were either above or below zero exceeded 95% (i.e., $p < 0.05$). To keep the statistical comparison transparent, for time intervals where significant differences were not theoretically expected (e.g. during the baseline period, or during the first cycle where REG and RND are indistinguishable), those identified clusters are presumed to be due to noise, were therefore marked as light grey in the figures below. The longest cluster identified in this way was used as a threshold for significance for the rest of the epoch, such that only clusters that exceed this duration are considered to be significant.

5.2.3. Tone Pattern Comparison Task (TP-COMP):

Following the EEG recording session, a behavioural task was carried out to evaluate the participants' deliberate auditory short-term/sensory memory ability (Schulze et al., 2011; Albouy et al., 2013; Graves et al., 2019). Participants were only informed of this task after the EEG session has concluded.

The task was similar to that reported in (Bianco and Chait, 2023) and Study 3: The stimuli contained two 500 ms tone-pip sequences separated by a 2000 ms silent gap (see Figure 5.2A). The sound sequences were comprised of ten 50 ms tone-pips drawn from the same pool described for the EEG stimuli above. Different patterns were drawn on each trial. The two sound sequences before and after the gap were matched on 50% of the trials ('same' trials) and differed in the other trials ('different' trials). The sequences in the 'different' trials were created by switching the positions of 3 of the 10 tones. The positions of the shuffled tones were randomly chosen on each trial, except for the first and last tones, to avoid primacy and recency effects. The instructions were to listen carefully to the sound sequences and press one of two keyboard buttons to indicated whether the two-tone sequences were the same or different ("S" for same and "D" for different). Participants then completed 32 trials. Feedback was provided after each trial. The correct response rate was used to quantify performance.

Before proceeding to the main task, subjects were given a practice task with 10 trials to ensure they understood the task structure. All sounds were generated anew for each subject in MATLAB at a 44.1 KHz sampling rate and stimulus delivery was controlled with psychotoolbox. Stimuli were delivered over Sennheiser HD558 headphones via a UA-33

sound card. The testing took place in the same booth as the EEG experiment. The task, including instructions, took approximately 5 min to complete.

5.2.3.1. Time Frequency Analysis

Time-frequency representations of EEG data during REG and RND conditions were derived using a wavelet-based approach, implemented via the FieldTrip toolbox. Continuous wavelet transforms were applied to each sound sequence evoked trial (see Figure 5.1A for stimuli example) across a frequency range of 1.5 - 30 Hz, with a frequency resolution of 0.1 Hz steps. A wavelet width of 5 was utilised to achieve an optimal balance between time and frequency resolution. The most responsive 10 EEG channels for each participant and all trials in each channel were included in this analysis. Padding was set to a minimum of 2 seconds to mitigate edge artifacts during the time-frequency transformation.

The analysis focused specifically on the activity evoked by the third cycle of tones within each sequence. Due to the applied padding, only the first nine tones of cycle 3, spanning from 11 seconds to 15.95 seconds relative to the onset of the first tone, were included. Time-frequency representations for the averaged tone-evoked activities across nine tones were baselined from -500 ms to 0 ms prior to each tone onset. Subsequently, the data were averaged across 10 selected EEG channels for each participant to refine the spectral outputs.

Statistical comparisons between two distinct groups of subjects—TP-COMP high performers (N=14) and low performers (N=16)—were conducted using FieldTrip's implemented function ft_freqstatistics. This analysis employed a Monte Carlo method coupled with an independent samples T-test to robustly determine significant differences between the groups across all assessed frequencies and time points. To ensure the reliability of the findings, 1000 random permutations were executed, and the significance threshold was established at an alpha level of 0.05, based on a two-tailed distribution.

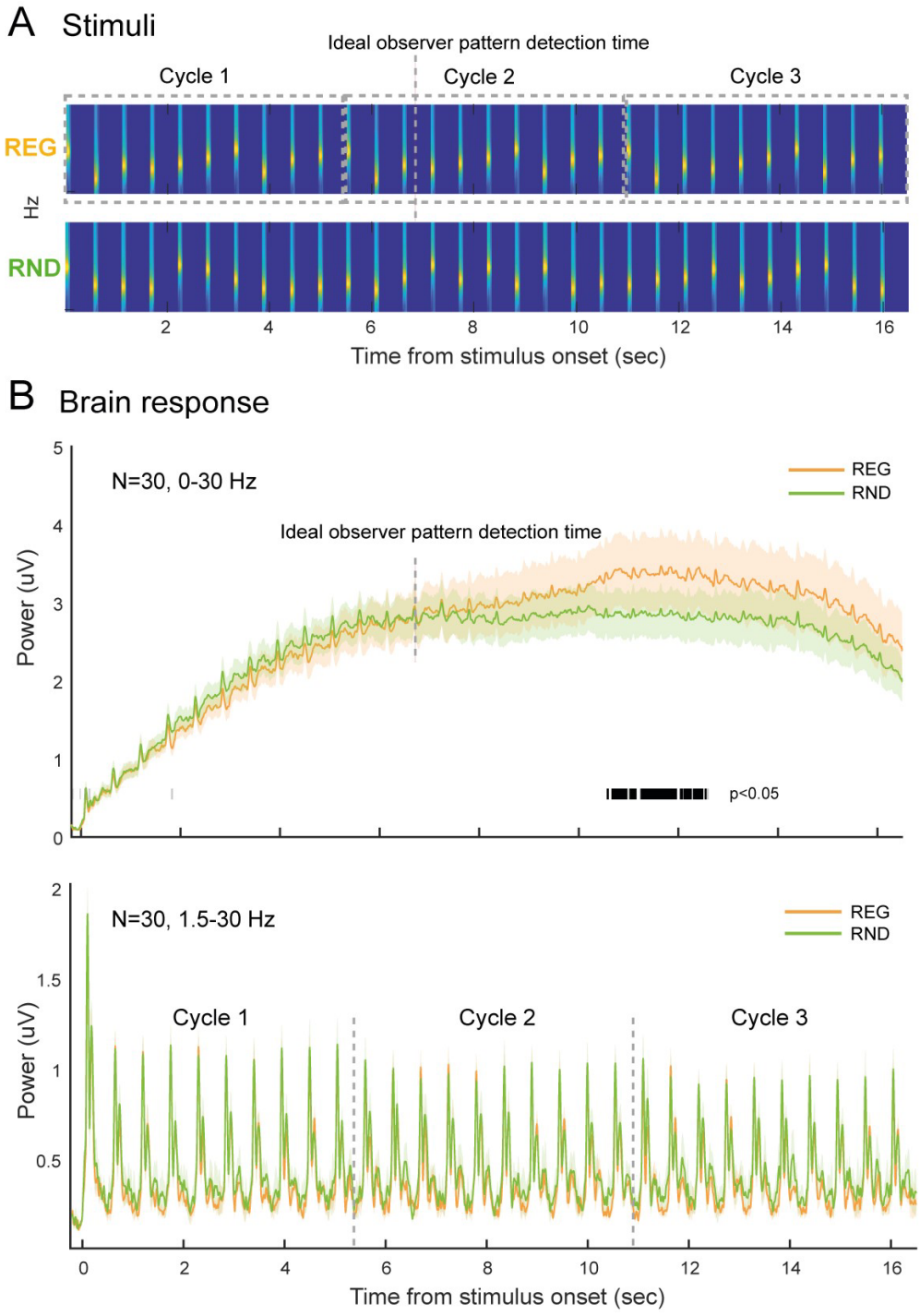

Figure 5.1. Brain response (A) Spectrogram of example REG and RND sequences used in the EEG experiment. All stimuli consisted of thirty 50 ms tones (3 regularity cycles in REG sequences; 16.5 s long), with 500ms silent gaps between tones. Naïve participants listened to the sound sequences passively while watching a movie. It is hypothesised that if the

brain monitors the transition probabilities between tones in the unfolding sequence, responses to REG and RND should be differentiated during the second cycle (when the pattern begins to repeat). Ideal observer-based estimates (e.g. Barascud et al, 2016) suggest an ideal observer requires roughly 3 or 4 tones (marked by the dashed line) after the onset of cycle 2. (B) EEG response evoked by REG and RND sequence. The top panel displays group RMS across a frequency range of 0-30 Hz for the full group of 30 participants (N=30). The two traces represent the average power during the presentation of regular (REG) versus random (RND) tone sequences. The shaded areas denote the standard error. The black bars below the graph indicate periods where the power differences between REG and RND conditions are statistically significant ($p<0.05$), suggesting that the regular patterns evoke a differentiable neural response at certain time points. The bottom graph in Panel A shows group RMS in a narrower frequency band of 1.5-30 Hz which emphasizes the tone-evoked activity (1.8 Hz).

5.3. Results

Neurophysiological (EEG) responses to auditory sound sequences were examined in thirty participants and related to short-term memory abilities probed with a Tone Pattern Comparison Task (TP-COMP).

5.3.1. Sequence-Evoked EEG Responses Suggest Regularity Extraction

The study explores brain responses to sound sequences, with a particular focus on the slow, sustained EEG response. The core question being investigated is whether the brain can automatically (outside of behavioural relevance) recognise slow REG patterns (cycle duration of 5500ms). This is being studied by testing whether brain responses to REG patterns differ from matched random sequences (RND). The upper graph of **Figure 5.1B** presents the group neural activity (mean and standard error of individual RMSs) to RND and REG sequences. A typical onset response is observed, which is then followed by a heightened level of sustained activity. This ongoing activity is punctuated by clear fluctuations at 1.8 Hz, mirroring the frequency at which the tones were presented. This observation is consistent with past MEG work, which has reported a similar pattern of sustained neural activity for patterns presented at 4 Hz (Hu et al., 2024).

Intriguely, a subtle enhancement in this sustained neural activity is observed when the brain is processing REG sequences in comparison to RND ones, particularly evident from the third cycle onwards. This suggests the brain exhibits sensitivity to patterns despite the effects is small.

5.3.2. Performance in Tone Pattern Comparison Task exhibit Significant Variability

The Tone Pattern Comparison Task (TP-COMP) serves as a classical method for evaluating participants' auditory short-term memory (Schulze et al., 2011; Albouy et al., 2013; Graves et al., 2019). Participants engage with this task by listening to pairs of tone sequences and determining whether each pair is same or different. The analysis reveals significant variability in participant performance, as depicted in **Figure 5.2**B.

The distribution of correct scores was consistent with a prior study (Bianco and Chait, 2023) and Study 3, with mean and median around 62% but quite a large variability across participants, likely reflecting variance in short term memory capacity. To further explore the relationship between task performance and brain activity, participants were divided into two groups based on the median performance score of the group distribution. This split resulted in one subgroup, 'Mem – high performers" displaying good ability in recognizing changed tone patterns, and another subgroup 'Mem – low performers" performing below the median, indicating less sensitivity to changes in the sequences.

5.3.3. TP-COMP Performance Associated with Differential EEG Responses to REG Patterns

The ability to monitor the structure of the sensory signal and identify the regularity is believed to be inherently intertwined with memory process. Therefore, this study explored the connection between participants' TP-COMP performance and the neural responses to sound sequences, as measured by EEG.

Remarkably, Mem – high performers demonstrated a pronounced enhancement in sustained neural responses to REG sequences, discernible from the second cycle of REG presentations, in contrast to RND sequences (illustrated in **Figure 5.3**A). This amplification in DC amplitude underscores a heightened neural activation to auditory regularities in high performers. Conversely, Mem – lower performers exhibited no significant differentiation in

their EEG responses to REG versus RND sequences, indicating a potential link between TP-COMP performance and the neural encoding of auditory patterns (as shown in **Figure 5.3**B).

Figure 5.3D consolidates the conclusion of interaction, presenting a comparative analysis of neural activity (REG-RND) between the two performance groups. Notably, significant clusters ($p<0.05$, bootstrap resampling) were identified from the third tone in the second REG cycle until the end of the sequence.

To further investigate potential correlations between sensory memory performance and sustained response dynamics across subjects, an exploratory Spearman correlation analysis between the difference in sustained response (REG vs RND) and the performance on TP-COMP was conducted. The correlations were conducted over 0.5 s intervals sampling the entire epoch (**Figure 5.3**D). The analysis revealed significant clusters ($p<0.05$), mainly within the 7.5-11 sec time interval, which corresponds to cycle 2 of REG – the initial period over which REG becomes theoretically detectable. The observation of correlation provide evidence that the link between sensory memory and the sustained response may be specific to this initial period where strong differences (DC started to diverge between RND and REG) are observed between conditions.

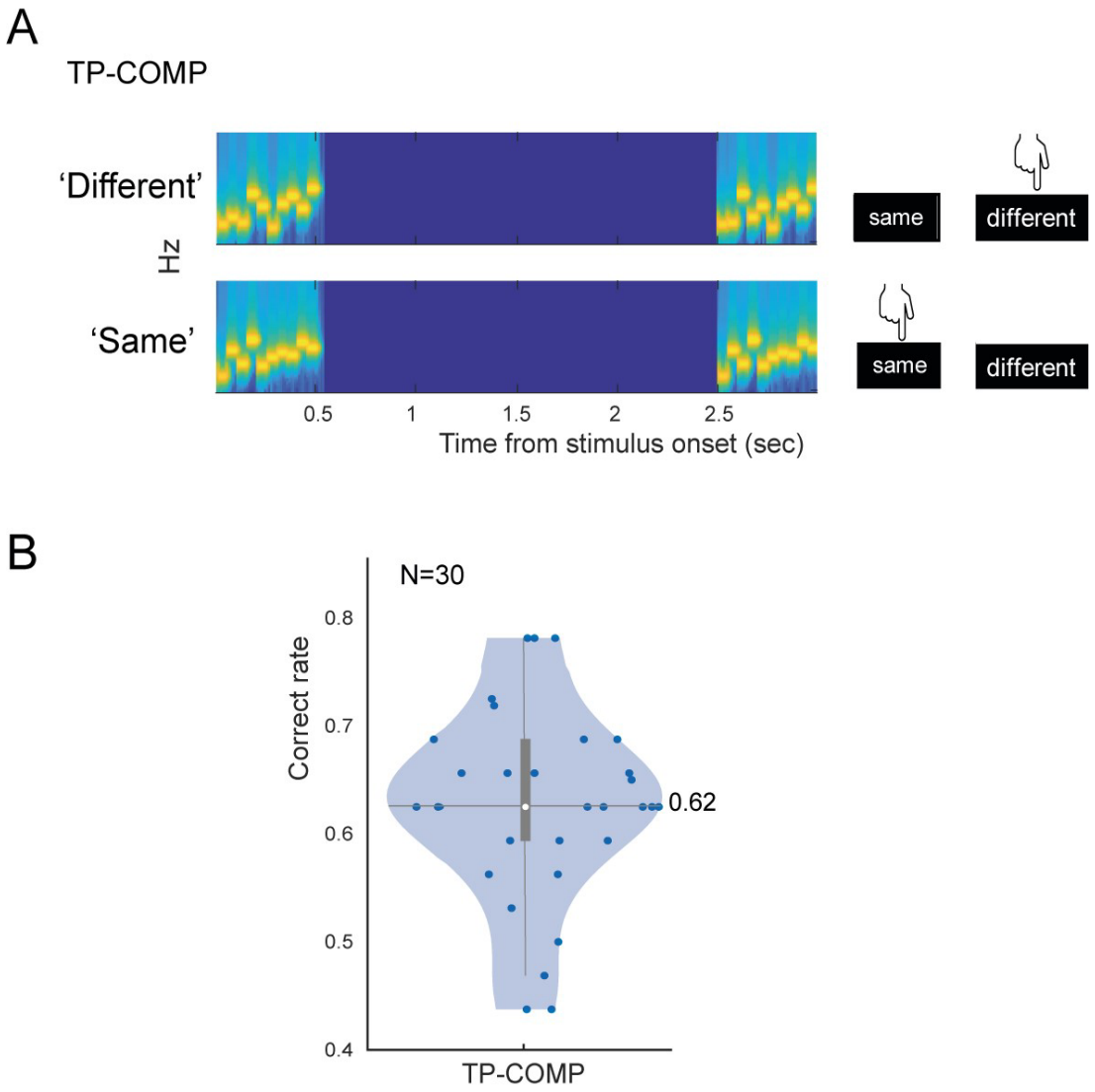

Figure 5.2. Memory task performance. (A) Examples of the stimuli used in the Tone Pattern Comparison Task (TP-COMP). Participants listened to two trials of a 500ms tone sequence, separated by a 2-second silent interval. They were then asked to indicate whether the two sequences were the same or different. In the 'different' trials, the positions of three tones were altered. However, the first and last tones were never changed to avoid primacy and recency effects. **(B)** Distribution of correct rates in the TP-COMP across all participants (N=30). The central dot at 0.62 indicates the median correct rate, while the thickness of the shaded area at different levels of correct rate suggests the density of participants scoring at that level. The plot shows variability in performance.

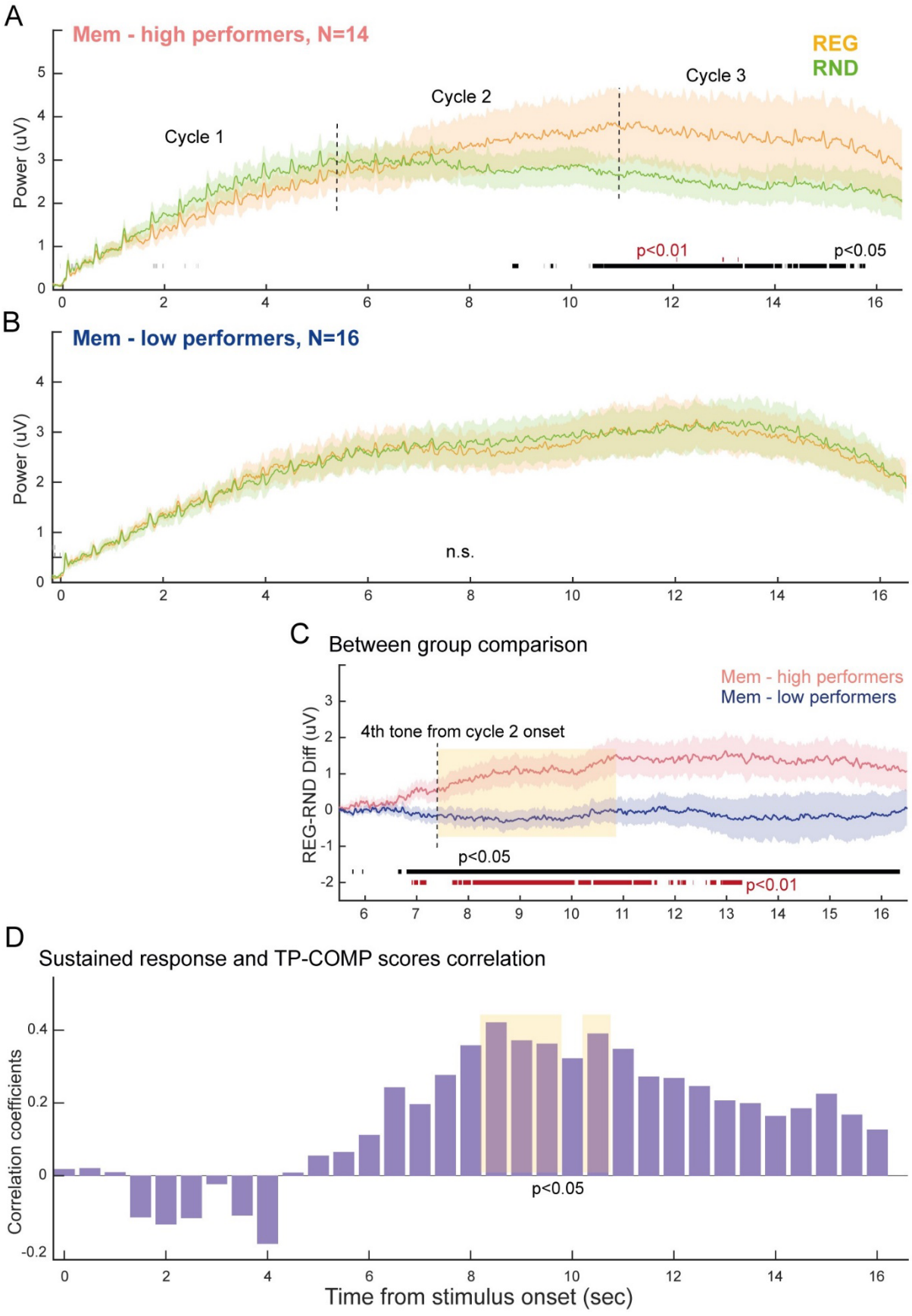
A
Mem - high performers, N=14
REG
RND
Cycle 1
Cycle 2
Cycle 3
Power (uV)
p<0.01
p<0.05
B
Mem - low performers, N=16
Power (uV)
n.s.
C
Between group comparison
Mem - high performers
Mem - low performers
4th tone from cycle 2 onset
REG-RND Diff (uV)
p<0.05
p<0.01
D
Sustained response and TP-COMP scores correlation
Correlation coefficients
p<0.05
Time from stimulus onset (sec)

Figure 5.3. Sequence evoked EEG responses grouped based on TP-COMP performance. (A) Group RMS mean over time for participants who scored above the median in the TP-COMP (N=14). The orange line indicates neural responses to regular (REG) tone sequences, while the green line shows responses to random (RND) sequences. The shaded area represents the standard error. Notably, there are power increases for the REG condition during cycles 2 and 3, as revealed by horizontal bar (black marks $p<0.05$, red marks $p<0.01$) at the bottom. This suggests enhanced stimulus processing in REG relative to RND. Note that the significant clusters observed during the pre-stimulus interval and first cycle, where no differences between REG and RND were expected, are labelled as grey. This applies to the clusters observed in the later time domain, where any identified clusters shorter than the longest cluster observed in those noise-attributed time intervals are also labelled as grey. (B) Group RMS mean over time for participants who scored below the median on the TP-COMP (N=16). Similar to panel A, two lines represent the average power for REG and RND sequences, with shaded areas indicating error. In this case, the lines overlap considerably, and no significant difference between REG and RND conditions was found. This suggests a less pronounced neural response to the stimulus' regularity. (C) Direct comparison between memory performance groups. The plot compares the neural responses differences of REG and RND between memory (TP-COMP) high performers and memory low performers. The memory high performers (shown in red) exhibit a significantly higher difference between REG and RND relative to the low performers. (D) The Spearman correlation between the difference in sustained response differences (REG vs RND) and correct score of TP-COMP was conducted with bin of 0.5 sec (2 Hz) over the entire trial duration. Each purple bar represents the Spearman correlation coefficients at each bin. Yellow shaded areas mark the time intervals where a significant correlation ($p < 0.05$; FWE uncorrected) was observed.

5.3.4. Tone-Evoked Responses Indicate Regularity Encoding

Following the sequence-evoked response, the phasic responses (1.5 – 30 Hz) elicited by tones presented in either regular (REG) or random (RND) sequences for all 30 participants were analysed. The extracted EEG data was divided into epochs, ranging from 50 milliseconds before to 500 milliseconds after each tone onset, centring the analysis on the exact moments when the brain processes each individual tone.

Figure 5.4 displays the group tone responses averaged across each REG cycle, and the corresponding ten tones of RND. The neural processing of individual tone is associated with four distinct peaks summarised by the root mean square (RMS) of the tone-evoked activity. The P1 component (40-60ms) appears at approximately 50 milliseconds after tone onset, indicating the brain's early sensory processing of the auditory signal. Following is the N1 component (80-120ms), peaking around 100 milliseconds, which is an early neural marker influenced by the predictability of auditory stimuli and highlights the brain's responsiveness to sensory signal (Todorovic and Lange, 2012; Hu et al., 2024). Subsequently, the P2 component (180-230ms), which arises close to 200 milliseconds post-tone, signals a deeper level of auditory processing that is hypothesised to be engaged in some aspect of the stimulus classification process (Crowley and Colrain, 2004). The final component, peaking around 330 milliseconds (250-400ms), appears to coincide temporally with N2 or N300 as reported in previous literatures (Näätänen, 1990; Renoult et al., 2012; Kumar et al., 2021a). However, no consistent hypotheses about the neural processes underlying this time interval exist, and it appears to be a component influenced by the stimulus context. For instance, the N2 component is noted for its sensitivity to various contexts, such as the detection of perceptual novelty or template mismatch (Crowley and Colrain, 2004). The topographic maps of the scalp voltage distribution for each neural component are shown in Figure 5.4. Interestingly, the N2 response exhibits a frontocentral distribution, with maximal negative amplitude observed at central midline electrodes, which exhibit similar topographic pattern as N1 (more discussion about N2 will be provided in the next section).

No differences were seen in cycle 2 where the evidence of regularity has been accumulating. However, in cycle 3, significantly increased power evoked by REG relative to RND are seen in P2; and more significantly reduced power evoked by REG compared to RND are seen in N1 and N2.

To evaluate whether and how REG and RND condition-specific tone-evoked responses change over time, the average evoked field differences between tones presented in the first and subsequent cycles were calculated. As the responses to the initial tones (the first two tones in Cycle#1) were influenced by onset-response activity, the last eight tones of each cycle (Cycle#1, tones 3-10; Cycle#2, tones 13–20; and Cycle#3, tones 23-30) were extracted and averaged. The average tone-evoked response during Cycle#1 (calculated across the time window of each neural component P1, N1, P2, and N2) were subtracted from that of Cycle#2 and Cycle#3 to assess how regularity modulates tone responses over the evolution of cycles. The results are displayed in Figure 5.4D.

For P1 (40-60ms), repeated measures ANOVA with sequence condition (RND or REG) and cycle number as factors found no significant effects of sequence condition ($F(1, 29) = 0.339$, $\eta2 = 0.012$, $p = 0.565$), cycle number

(F(1, 29) = 0.708, η2 = 0.024, p = 0.407), or their interaction (F(1, 29) = 1.924, η2 = 0.062, p = 0.176).

For P2 (180-230ms), there were no significant effects of sequence condition (F(1, 29) = 2.231, η2 = 0.074, p = 0.138), cycle number (F(1, 29) = 2.987, η2 = 0.064, p = 0.169), or their interaction (F(1, 29) = .056, η2 = 0.002, p = 0.815).

For N1, a significant effect of cycle number was observed (F(1, 29) = 19.616, η2 = 0.403, p <.001) where the cycle#3-cycle#1 exhibited reduced amplitude compared to cycle#2-cycle#1. However, there were no significant effects of sequence condition (*F*(1, 29) = 0.355, *η*2 = 0.012, *p* = 0.556) or the interaction of those two factors (F(1, 29) = 2.104, η2 = 0.071, p = 0.148).

For N2 (180-230ms), there were no significant effects of sequence condition (F(1, 29) = 0.03, η2 = 0.001, p = 0.869), cycle number (F(1, 29) = 0.826, η2 = 0.028, p = 0.371), or their interaction (F(1, 29) = 2.168, η2 = 0.07, p = 0.152).

The reduced N1 amplitude between cycle#3-cycle#1 and cycle#2-cycle#1 was observed in both RND and REG conditions, as shown by ANOVA. Significant differences in N1 for cycle#2-cycle#1 and cycle#3-cycle#1 in the RND condition were further confirmed by a one-sample t test. This test compared the values to zero, indicating a consistent reduction relative to cycle 1 [Cycle#2: t(1,29) = −3.1088, d = −0.5679, p = 0.0042; Cycle#3: t(1,29) = −5.4675, d = −0.9982, p <0.001]. The same test for the REG condition also showed significant effects [Cycle#2 t(1,29) = −4.5209, d = −0.8254, p <0.001; Cycle#3 t(1,29) = −8.5236, d = −1.5562, p <0.0001]. This finding contrasts with the MEG findings that identified differences distinctively in the regular (REG) condition (Hu et al. 2024) and did not detect significant impact of cycle positions on N1 responses in both regular and random conditions. However, this experiment reveals an unexpected pattern, where N1 amplitudes is reducing over cycles.

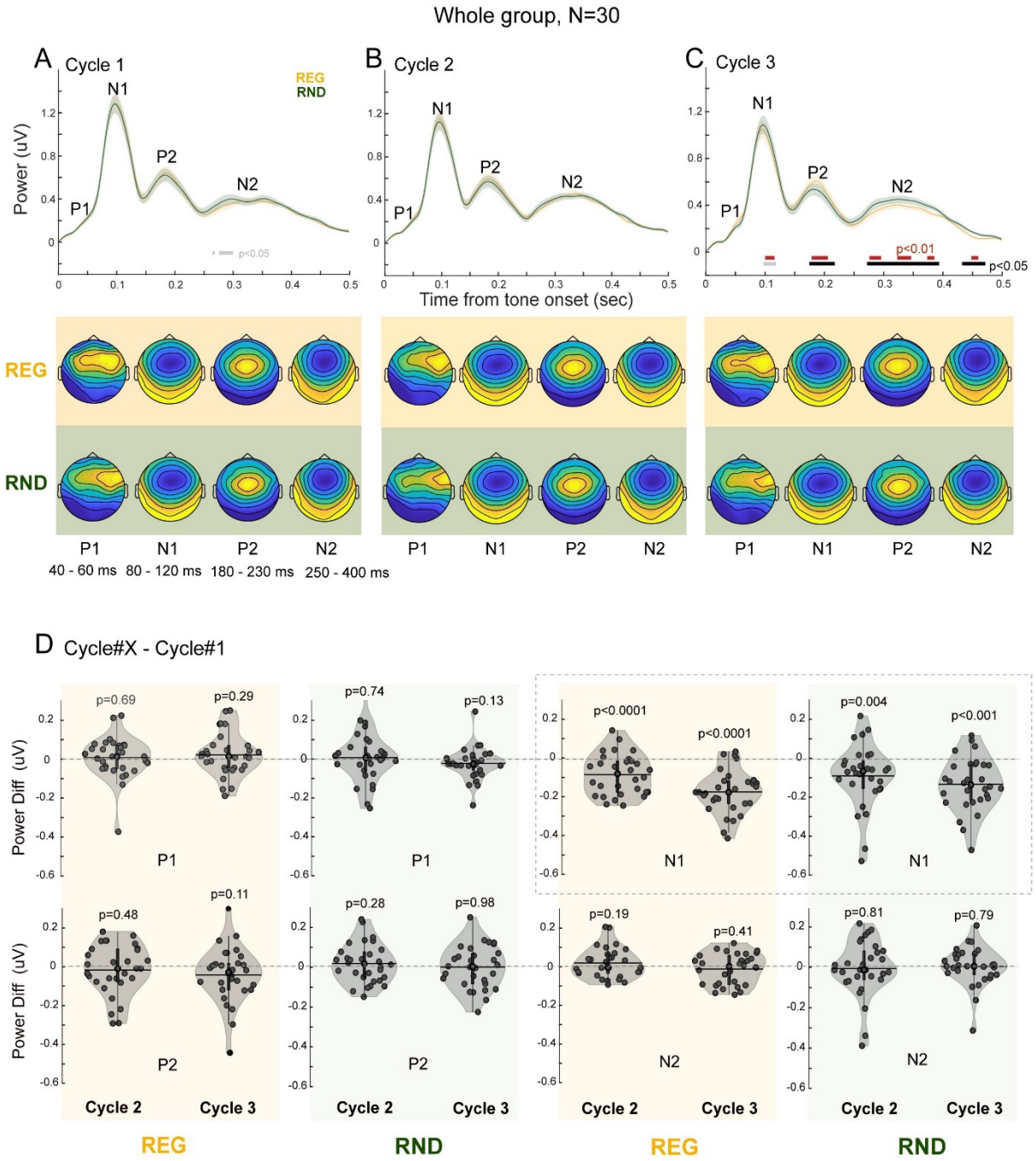

Figure 5.4. Tone response across each cycle of REG and corresponding timing in RND. (A) Tone evoked activity in cycle 1. Four distinctive ERP components are visible, P1 (40-60ms), N1 (80-120ms), P2 (180-230ms), N2 (250-400ms). Any differences between conditions here (indicated in light grey) are considered noise and used as a threshold of significance in cycle 2 and cycle3. (B) Tone evoked activity in cycle 2. The Regularity (REG)

started to establish in this cycle, however, no significant differences were seen between conditions in any ERP components. (C) Tone evoked activity in cycle 3. Regularity has been theoretically established in this cycle. Significant differences between conditions were seen in different ERP components. The black and red horizontal bar represents the statistical significance. The grey horizontal bar indicated intervals that did not pass the cluster threshold. The topographical maps represent the scalp voltages averaged across each time interval of interest. (D) To test how the mean amplitude of neural components changes over the course of cycles, the difference from the first cycle was calculated over the time interval of P1 (40-60ms), N1 (80-120ms), P2 (180-230ms), and N2 (250-400ms) for the second and third cycles in REG and RND. Only the last eight tones in each cycle were included in this analysis to avoid the onset response in the first cycle. Statistical differences between the cycle differences and zero were tested using a one-sample t-test in each condition, and the p-value was labelled on top of each distribution. Significantly reduced responses were observed only in N1 with both RND and REG. Specifically, REG shows a stronger magnitude of reduction over the course of cycles compared to RND in N1 time window.

5.3.5. Inter-Group Comparisons of Phasic Responses Highlight Variations in Auditory Processing of Individuals

Similar to the sequence-evoked response analysis, the relationship between neural responses to individual tones and short-term memory performance were examined. For this, participants were categorised based on their performance on the TP-COMP similarly as sequence response analysis (**Figure 5.5**).

Figure 5.5A displays the averaged tone-evoked activity in each cycle of REG and corresponding time interval of RND. The low performer group did not show any significant effects, but significant clusters were observed for the high performers, as indicated in Figure 5.5A.

Figure 5.5B shows the interaction between participants' group categories (high and low performers, based on TP-COMP scores) and the type of sequence (random or regular). Here, significant clusters were detected exclusively within the N2 component's timeframe (250ms - 400ms), highlighting a link between the N2 component and short-term memory

capabilities. The independent t test examined the group differences between mean amplitude over the N1and N2 time window, confirming the significantly reduced amplitude [t(1,29)=-2.6739, d=-0.9785, p=0.0124] of N2 in Mem-high performers, compared to low performers. No significance was seen in N1 [t(1,29)=-1.1836, d=-0.4331, p=0.2465].

To summarize, for the tone-response level, a whole group analysis suggested potential differences between REG and RND during the N1, (RND>REG; consistent with Hu et al, (2024)), P2 (REG>RND; but this is difficult to interpret due to baselining, and the increase in P2 typically coincides with the reduction in N1) and N2 (RND>REG). A comparison between the low- and high- memory performers also revealed a specific difference between groups in the N2 range, with high memory performers exhibiting significant differences between conditions in that interval exclusively.

To investigate whether the observed relationships extend to individual differences, a Spearman correlation analysis was conducted. This analysis examined the relationship between TP-COMP scores and neural activity throughout the entire tone epoch - baseline-corrected at tone onset - across participants (refer to **Figure 5.5**C). Interestingly, negative correlations were noted between TP-COMP performance and the N2 time interval, suggesting the better short-term memory ability is associated with the larger differences between RND and REG. Notably, within the N2, statistical significance was only observed within the 0.25-0.3 sec post-tone onset across subjects. In addition to the potential for false positives, these findings suggested that only specific neural processes underlying the tone-evoked response are correlated between short-term memory ability and regularity encoding.

Finally, the neural oscillations that could potentially explain the observed group differences in the time domain were analysed (**Figure 5.5**B). The comparison of time-frequency domain for tone-evoked activity in cycle 3 between TP-COMP high performers and TP-COMP low performers is illustrated in **Figure 5.5**D. Significant clusters of activity are observed across multiple event-related potential components, namely P1, N1, P2, and N2, manifesting as a broadband pattern. Notably, slow theta-like oscillations were initially observed from tone onset until approximately 250 ms, subsequently re-emerging between 330 ms and 410 ms. During the N2 time interval, where persistent and significant differences between the two groups were observed (see **Figure 5.5**B), the time-frequency representation reveals a broadband encompassing theta, alpha, and low beta frequencies. These findings suggested that participants who perform above the median on the TP-COMP task exhibit reduced brain activations in response to tones within a REG sequence in cycle 3, compared to a RND sequence. This reduction in neural activity spans multiple frequency bands, indicating that the observed differences during the N2 interval may be attributable to multiple neural processes. Interestingly, those results align with the recent MEG study which suggests that increased alpha and beta power are associated with disruption of learned music sequences (Bonetti et al., 2024).

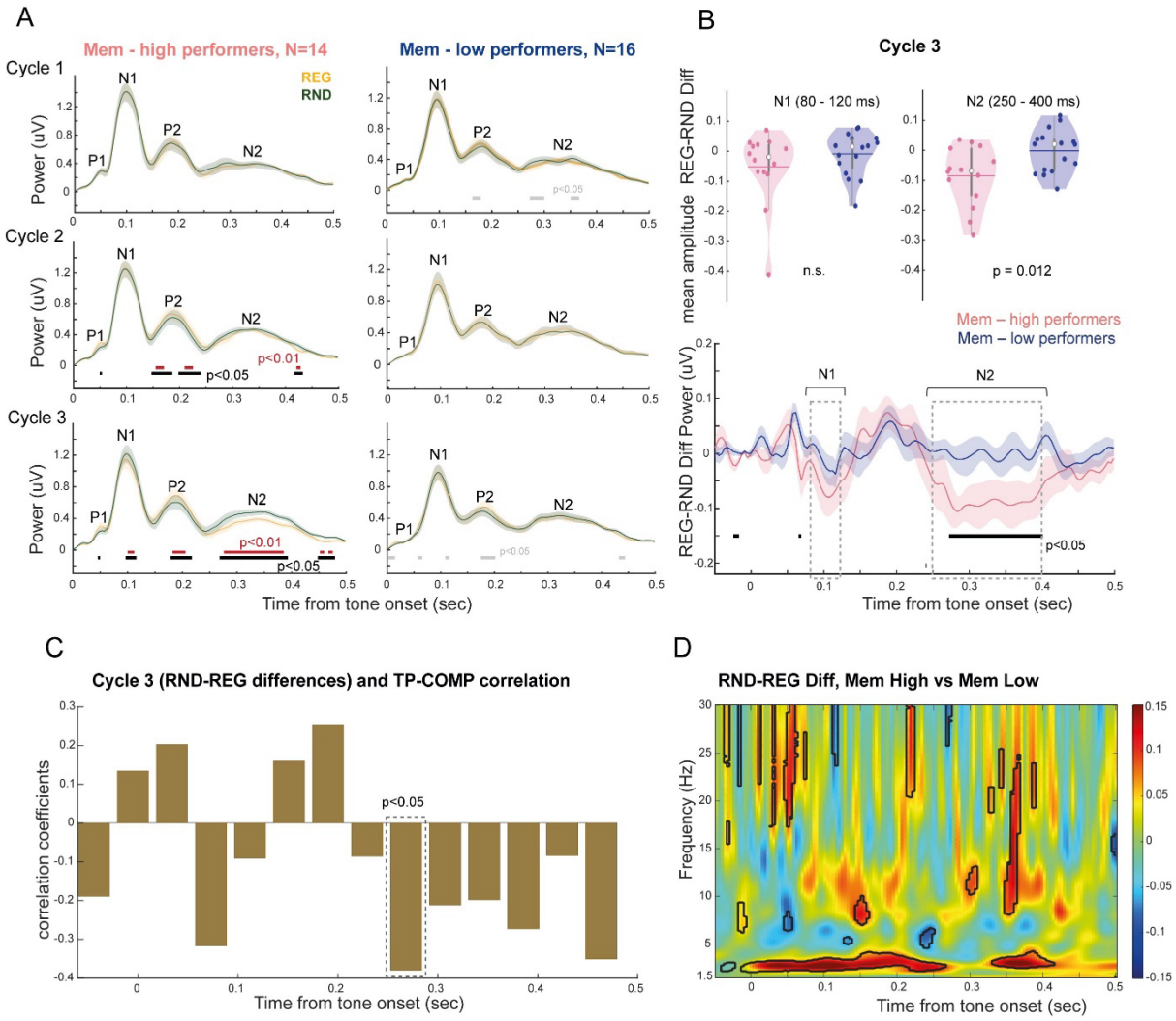

Figure 5.5. Tone response comparison between TPCT above median performers and below median performers. (A) Tone evoked response. Left column shows tone responses of memory high performers in TP-COMP (N=14) and the right shows those memory low performers (N=16). One row for each cycle, demonstrating the power averaged across tones in each cycle. Red and black horizontal bar represents the significant cluster. Grey bar represents the noise-attributed clusters (see methods). (B) Between group comparison. The bottom graph shows the tone response difference (REG-RND) in each group. The upper graphs represent the mean response distribution of each time window of interest across participants in each group. Statistical comparison suggests a

significant difference in N2 (250-400ms) between memory high performers and low performers; that those memory high performers exhibit a reduced N2 response in REG relative to RND. (C) Correlation analysis. The Spearman correlation between the difference in tone response (REG-RND, cycle 3 of REG and corresponding time interval of RND, baselined from tone onset) and correct score of TP-COMP was conducted with bin of 40 ms over the tone epoch across all subjects (N=30). Each brown bar represents the Spearman correlation coefficients summarised within each bin. Dashed square mark the time intervals exhibit significant effects ($p < 0.05$; FWE uncorrected). Negative correlation coefficients with statistical significance are observed within the time interval of N2, specifically between 0.25-0.3 sec. (D) Time frequency domain of tone response in cycle 3. The RND-REG differences comparison between TP-COMP high performers and low performers. Warm colour represents stronger RND-REG power differences. Cold colour represents stronger REG-RND power differences. The black line marks the clusters that exhibit significance ($p < 0.05$).

5.3.6. Enhanced Early Auditory Response was Observed in Memory Task High Performers

This study further examined whether individuals' baseline auditory responses (collapsed across REG and RND evoked tone responses), differ based on their performance on the TP-COMP, as shown in **Figure 5.6**. Analysing the tone responses from the first cycle across both REG and RND conditions, it was found that participants with TP-COMP scores above the median (Mem – high performers) exhibited significantly enhanced P1 and N1 responses, potentially indicating more efficient early sensory encoding of auditory stimuli (Näätänen and Picton, 1987; Stufflebeam et al., 1998). Interestingly, this enhanced encoding was not evident in later neural components like P2 (Crowley and Colrain, 2004) or N2 (Folstein and Van Petten, 2008).

To investigate the relationship between neural responses and behavioural performance across subjects, the Spearman correlation analyses between TP-COMP scores and P1 and N1 amplitudes was conducted. However, as suggested by **Figure 5.6**B's scatter plots, no significant correlations for either P1 ($rho=0.22$, $p=0.17$) or N1 ($rho = 0.29$, $p=0.11$) were observed. The lack of significant correlation between P1, N1 amplitudes and TP-COMP scores suggests that those early sensory processing biomarkers are not the solely predictors of individual performance on TP-COMP. The observed correlation across

subgroups might indicate that participants who possess better sensory encoding precision tend to detect the tone pattern change better. However, the ability to detect the change also requires the ability of maintaining the complex tone pattern and retrieving the information when necessary. This could be attributed to the task's dependence on neural processes such as memory, which extend beyond the initial sensory encoding.

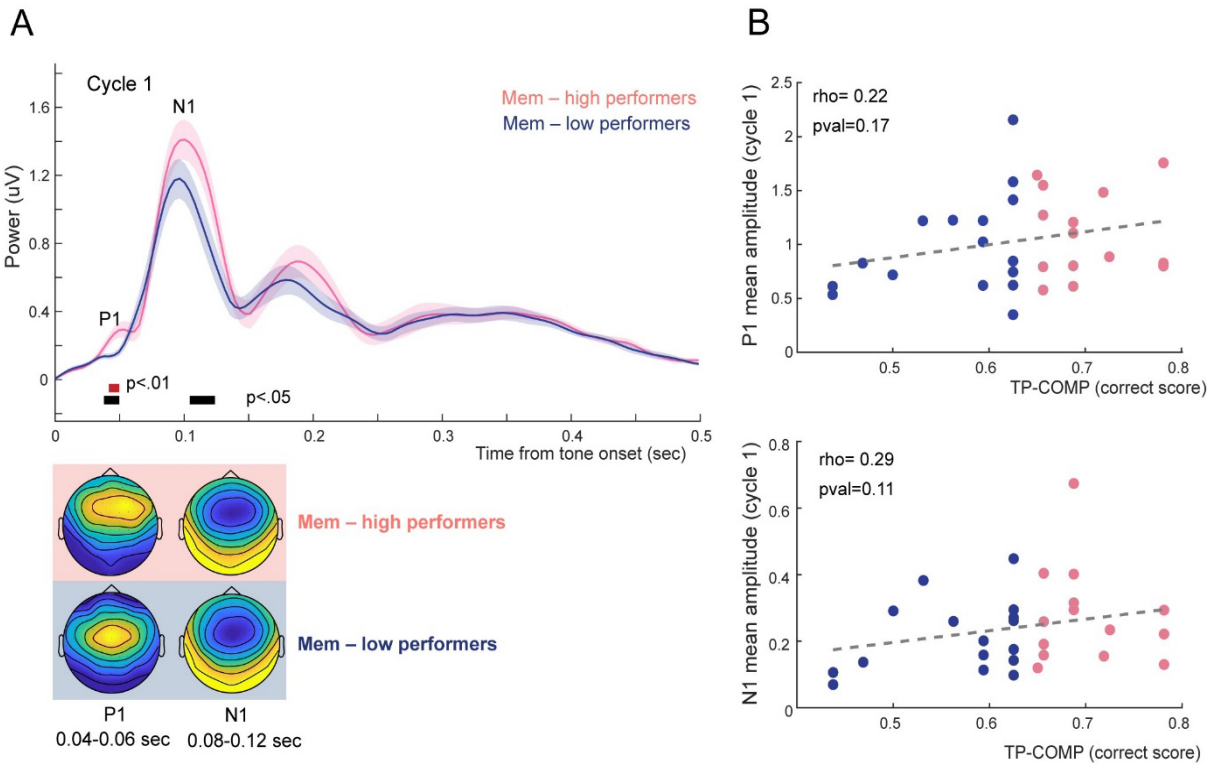

Figure 5.6. Tone-evoked activity in cycle 1. (A) Tone responses averaged across the first cycle for both RND and REG in each group of performers. The blue line represents high performers in the memory task (TP-COMP), and the pink line represents low performers. Statistically significant differences between groups were observed in the P1 (40-60ms) and N1 (80-120ms) components. (B) Scatter plots for neural responses in P1 (0.04-0.06 s), N1 (0.08-0.12 s) averaged across cycle 1 and TP-COMP scores. Spearman correlation between correct scores on the memory task and mean amplitudes for P1 and N1 components are provided in each plot. The aim is to test whether these two ERP components correlate with memory task performance as suggested by Figure A. However, the upper plot illustrates a non-significant correlation for the P1 amplitude (rho = 0.22, p = 0.17), and the lower plot shows a

slightly stronger, yet still non-significant correlation for the N1 amplitude (rho = 0.29, p = 0.11).

5.3.7. Correlation Between Tone-evoked Activity and Sustained Response

In this study, the analysis reveals that the N2 response (250-400ms) which is hypothesised to be associated with more advanced neural processing (Näätänen, 1990), is significantly correlated with memory performance revealed by TP-COMP (see **Figure 5.5**C). Interestingly, as indicated by earlier analysis, a significant correlation is also observed between the amplitude of sustained responses and memory performance across participants, particularly in 7.5-11 sec of cycle 2 (**Figure 5.3**C,D). It is important to mention that the amplitude of sustained responses is hypothesised to reflect the precision coding of sensory signals (Zhao et al., 2024). Therefore, to further explore the potential link between tone-evoked activity (N1, N2) and sequence-evoked activity (sustained response), the mean amplitude of the sustained response and tone responses (N1 and N2) were extracted and correlated. Specifically, the Spearman correlation was conducted on the REG-RND difference in the tone-evoked response (high-pass filtered, see **Figure 5.4**), and the REG-RND difference in the sustained response (low-pass filtered, as shown in **Figure 5.1**B) during Cycle#2 (7.5-11 s, when the response to REG began to diverge from RND, see **Figure 5.1**B) and Cycle#3 (11-16.5 s), across all participants.

The results presented in **Figure 5.7** indicate a moderate but significant negative correlation between the REG-RND difference in sustained responses and those of N1 response (rho = -0.40, p = 0.02) in Cycle 2. However, this significance value did not survive the p value threshold after the Bonferroni correction (p value threshold = 0.0125). Meanwhile, no significant effects are observed between the N2 response and the sustained response in the same timeframe (rho = -0.17, p = 0.36).

Interestingly, in Cycle 3, when the pattern has been theoretically established, the correlation between the N1 response and sustained response does not reach statistical significance (rho = -0.27, p = 0.14). Conversely, a significant correlation is seen between the N2 response and the sustained response (rho = -0.45, p = 0.012), and the significance value also survived the Bonferroni correction, providing strong evidence that the more reduced responses of N2 in REG compared to RND are associated with the increased enhancement in sustained response in REG relative to RND in Cycle 3.

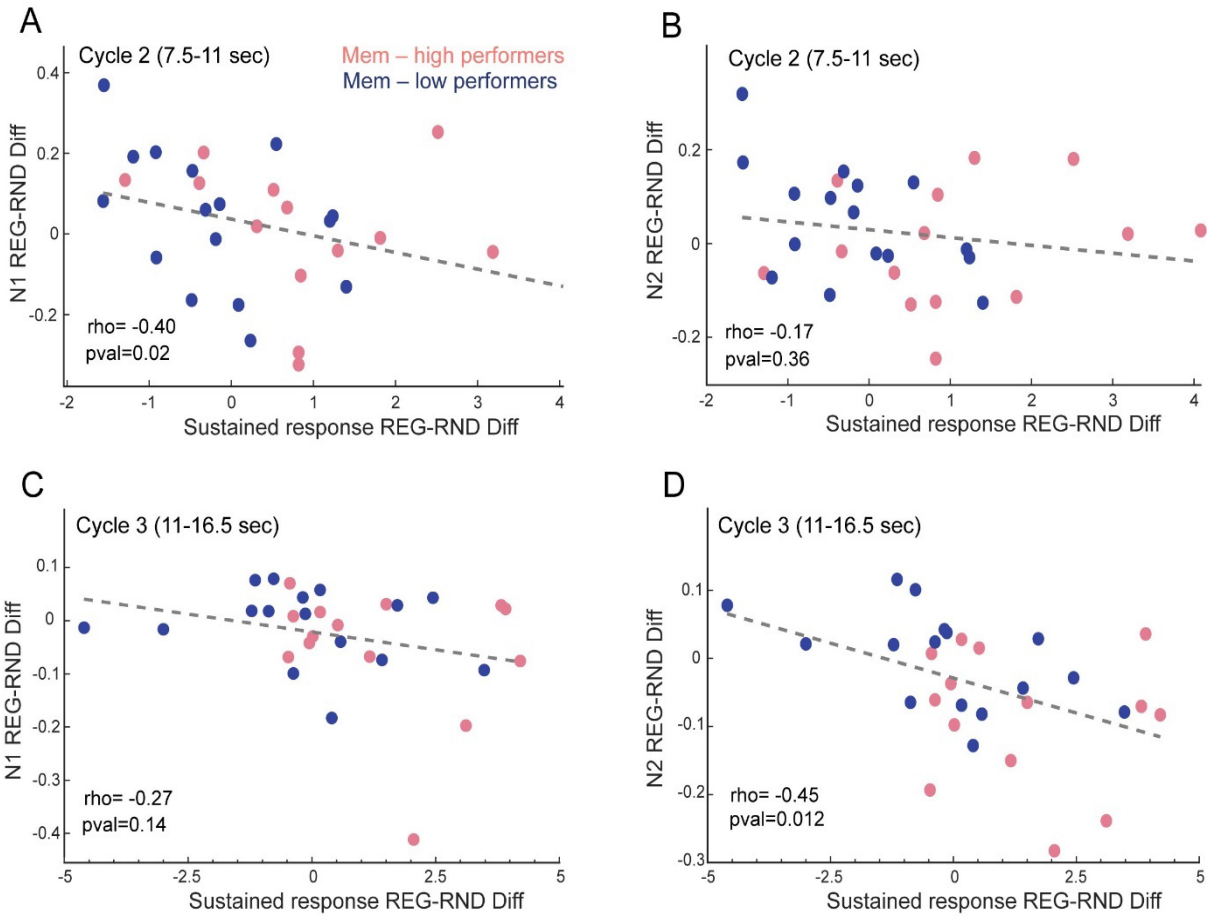

Figure 5.7. Scatter plots for neural responses in the 7.5-11 s (Cycle 2) and the 11-16.5 s (Cycle 3) time windows. Spearman correlation between mean amplitude of sustained response differences (REG-RND) and REG-RND differences of N1 (0.08-0.12 s), N2 (0.25-0.4 s) are provided in each plot. Blue dots represent high performers in TP-COMP, pink dots represent low performers. (A) The correlation of neural activity, averaged across the time latency of the DC shift (where the divergence between REG and RND occurs) to the end of cycle 2 (see Figure 5.1B), shows a moderate negative correlation (rho=-0.4, p=0.02, uncorrected). This suggests that more reduced REG response relative to RND in N1 amplitude are associated with increased enhancement of REG response relative to RND in the sustained response across subjects. (B) No significant correlation was observed in Cycle 2 for N2. (C) Similar analysis

was conducted on time interval of Cycle 3; no significant correlation was seen for N1. (D) A significant negative correlation was observed for N2 (rho=-0.45, p=0.012, uncorrected) in cycle 3. This suggests that as the sustained response of REG relative to RND increases, the N2 amplitude of REG relative to RND decreases.

5.4. Discussion

In this study, electroencephalography (EEG) was utilised to measure the brain activity in human participants. The participants were engaged in passive listening to sound sequences that were either predictable or unpredictable. Two types of neural responses were investigated: sustained responses, to assess sensitivity to sequence regularity, and phasic responses, to evaluate event-evoked activity for individual tones. Following the EEG recording, participants were administered with the Tone Pattern Comparison Task (TP-COMP) to evaluate their auditory short-term memory capabilities.

The findings revealed a pronounced increase in the power of sustained responses upon the establishment of a regular pattern. Notably, the magnitude of this effect was modulated by the TP-COMP performance. Participants with higher task scores demonstrated a greater brain response distinction between the regular patterns (REG) and random sequences (RND), while the lower scorers exhibited negligible differences. This pattern of response also extended to phasic activity during the N2 time window following tone onset, with the higher TP-COMP scores correlating with a greater separation in N2 responses. Furthermore, the correlation analysis indicated that the differences in sustained responses between REG and RND conditions were inversely related to differences in N2 responses, suggesting that the two neural components might share the common underlying mechanisms, or be influenced by the same neural pathways.

5.4.1. Shared Neural Processes Between Implicit Sound Pattern Detection and Auditory Short-term Memory

This study explored the relationship between the sustained neural responses evoked by sound patterns and individual performance in the tone pattern comparison task (TP-COMP). The task revealed that the higher-performing individuals exhibit more pronounced sustained responses to sound patterns compared to their lower-performing counterparts. As previously discussed, the brain's ability to recognise sound patterns relies on maintaining a record of past auditory sequences so that it can be used to compare with

the incoming sensory inputs. This process has been found to be facilitated by the sustained neural responses within a network comprising the auditory cortex, inferior frontal cortex, and hippocampus (Barascud et al., 2016; Hu et al., 2024). In specific, hippocampus is known to be the primary hub for memory function in the brain, playing a role in both long-term and short-term memory (Kumaran, 2008), it is therefore hypothesised that this correlation might be associated with the engagement of hippocampus.

To support the above, relevant scenario such as research conducted by Kumar et al, (2014), demonstrated that human listeners could discriminate repetitive noise patterns through unsupervised learning upon prolonged auditory exposure. Using fMRI and multi-voxel pattern analysis, the authors found that hippocampus played an essential role in encoding long-term auditory experiences, which supports the noise pattern recognition (Kumar et al., 2014).

Indeed, hippocampus has been extensively reported across various studies related to working/short-term memory (Kumar et al., 2016, 2021b; Tsetsenis et al., 2023). However, it is important to note that the behaviour outcome measured by the TP-COMP task is accounted for different cognitive procedures, each potentially engaging specific brain regions crucial for supporting these processes. In other words, the correlation observed between sustained response and behaviour performance in current study could be explained by any stage assessed by the task.

To clarify the brain areas which are associated with the two stimulus comparison task, some previous studies have attempted to examine the neural involvement at each stage (Kaiser, 2015; Quentin et al., 2019; Yuan, 2019). For example, empirical evidence from fMRI data (Kumar et al., 2016) showed that during the encoding phase, listeners who are exposed to auditory stimulus (i.e. tone) would engage the auditory cortex. The maintenance phase, which involves retaining stimuli in memory, is found to be facilitated by sustained activation in both the auditory cortex and the hippocampus. Additionally, the study suggested that the inferior frontal gyrus supported the stabilisation during this phase, which is enhanced by its functional connectivity with the hippocampus. During the retrieval phase, participants compare a probe stimulus with the memorised one, significantly engaging both the hippocampus and frontal regions.

Furthermore, Kumar and colleagues explored the neural correlates of this auditory working memory. They tested the same paradigm on patients with medically implanted electrodes in regions related to auditory working memory. The analysis of local field potentials (LFPs) confirmed the sustained neural activity patterns in various brain regions during the maintenance phase of auditory working memory (AWM). The primary auditory cortex exhibited an increase in delta power along with a general suppression of higher frequencies, specifically in the beta and gamma ranges. Concurrently, the medial temporal

lobe regions, including the hippocampus and parahippocampal gyrus, demonstrated an enhancement in low-frequency activities, predominantly in the delta and theta bands, coupled with a reduction in high-frequency oscillations (Kumar et al., 2021b). Following that, it is plausible to hypothesise that the enhanced sustained responses observed in this study are likely linked to neural sources sharing mechanisms with low frequency oscillations such as delta, as suggested by Kumar's findings.

However, it is important to emphasize that this study did not demonstrate a significant correlation over the full trial of the sequence evoked activity across subjects (see Figure 5.3D), instead the above significant effects were only observed within cycle 2, where the sustained response evoked by the REG started to diverge from the RND. Specifically, this is the point when the neurons effectively detected the transition, hence the neural process underly the REG and RND started to differentiate. One explanation is that the signal to noise ratio is relatively higher in this dynamic process. Alternatively, this might suggest the neural circuits, responsible for signalling the pattern emergence (i.e. as indicated by the DC shift), are specifically interacted with the neural pathways of the TP-COMP.

Overall, as the TP-COMP performance variability may be accounted by multiple cognitive processes, future research should aim to measure the neural responses to both tasks and examine their temporal and spatial relationships, in order to further elucidate these complex interactions.

5.4.2. Neural Processes Underlying N2 are Associated with Regularity Encoding

Introducing silent gaps between successive tones has allowed to dissociate the neural responses that phase locked to each individual tone. Interestingly, the analysis in this study revealed a significant reduction in the N2/N300 time interval (250ms - 400ms post-tone onset) of phasic activity (1.5-30 Hz) for tones that consist regular pattern compared to a random sequence (see Figure 5.4). This suggested that the neural processes underlying this specific time interval exhibit sensitivity to the auditory sequential regularity (auditory context). Although this negative deflection have not been widely documented in past auditory literature, several studies have provided insights related to their association with the processing of unexpected or deviant auditory events compared to a standard input (Näätänen, 1990).

One relevant study by Besson et al. (2007) investigated how musical training enhances pitch perception within music, revealing that musical training sharpens the ability to detect pitch variations (Besson et al., 2007). The EEG data suggested that musicians tend

to exhibit more pronounced responses to subtle pitch changes compared to non-musicians, as evidenced by the N300 responses, which peak at 300ms after the sound onset, to pitch incongruities. This study demonstrated that the N300 component is crucial for illustrating auditory acuity that allows for the detection of subtle discrepancies in pitch, suggesting its sensitivity to the statistics of auditory context. Interestingly, the team also found that the response is particularly enhanced in musicians, an indicative that music training can significantly improve auditory processing. Their observations are further illuminated by evidence showing a correlation between musical proficiency and performance in short-term memory tasks (Fujioka et al., 2006; Parbery-Clark et al., 2009; Chandrasekaran and Kraus, 2010). In addition, musician's enhanced sensitivity to stimuli statistics for musicians was also reported from the previous study (Shook et al., 2013). Furthermore, the N300 component is featured in a more recent work by Randeniya et al. (2022). In their experiment, participants were exposed to a stochastic oddball paradigm while listening to sounds of varying frequencies, and the purpose was to test the brain's response to deviations from standard embedded within stochastic auditory patterns. The EEG data analysis revealed that the N300 time window is sensitive to the auditory context, with stronger responses to uncertain (deviant) sound relative to the certain (standard) sound. This observation led researchers to interpret this neural component as possibly representing the prediction error signal, highlighting its role in auditory context processing (Randeniya et al., 2022).

Although the N2/N300 component was not commonly reported in auditory research, it has been extensively explored within the field of vision and shares similar characteristics as the component discovered in this study (McPherson and Holcomb, 1999; Schendan and Kutas, 2002, 2003, 2007). Crucially, one insightful research utilised the predictive coding theory to explore the nature of N300. Through a series of experiments focusing on how the N300 is modulated by statistically regular and irregular visual scenes, researchers discovered that the amplitude of the N300 is significantly reduced when responding to regular scenes compared to irregular ones. This finding suggested that the N300 is sensitive to the statistical properties of visual inputs, indicating its role in processing prediction errors within visual information (Kumar et al., 2021a).

Since the N2's timing is relatively late compared to phase-locked responses observed in learning based on local statistics, such as stimuli occurrences and transitional probabilities (Todorovic and Lange, 2012; Maheu et al., 2019). Thus, this component appears to be less likely to stem from low-level bottom-up computations. As inspired by some previous work in visual modality (McPherson and Holcomb, 1999; Schendan and Kutas, 2002, 2003, 2007), it is plausible that neural processes underlying N2 are associated with high-level top-down expectations, for example, such as learning/representing the abstraction/summary statistics of the sound sequence. Correspondingly, the attenuated N2 response observed in the third REG cycle, where the auditory pattern was theoretically

established, supports the notion that the brain had accumulated sufficient evidence to compute stable statistics of the auditory stream (i.e. the precision of the auditory input). This likely enables the brain to suppress certain neural processes, thereby conserving resources to manage unexpected future scenarios.

Moreover, in terms of neuroanatomy, the 'late' nature of N2 suggests that it is less likely to originate from the primary auditory cortex. Instead, candidate regions might be posterior areas such as the planum temporale, which known for encoding abstract sound categories (Giordano et al., 2013), or frontal cortex, which is thought to be associated with sequential structure encoding (Stiso et al., 2022). Nevertheless, the current data, limited to only two predictability conditions, do not fully substantiate this hypothesis. The lack of spatial resolution in the 64-channel EEG cannot provide reliable evidence about where this component is sourced. Future research could expand on this by incorporating invasive tools and manipulating the predictability of auditory sequences more comprehensively to investigate the nature of this neural process.

5.4.3. Correlation Between Sequence Evoked Sustained Response and Tone Evoked Phasic Activity

Except for the N2's sensitivity to regularity, this study observed that N2 amplitude is also modulated by TP-COMP performance. High performers exhibit a significantly reduced response to REG relative to RND, compared to low performers. The correlation between TP-COMP task performance and N2, along with sustained response, implies shared or interacted neural circuits in these cognitive processes. In an exploratory analysis, the relationship between two neural components, both shown to be sensitive to stimulus predictability, was examined. The results suggested that the response differences between the REG and RND for N2 are inversely related to the sustained responses REG-RND differences in the last cycle of REG and the last 10 tones of RND. This may indicate that these manifestations are influenced by the shared neural pathways. To support this, similar insights were also suggested by the EEG study in a clinical context. Coffman and colleagues investigated the electrophysiological mechanism of how individuals with schizophrenia and healthy controls process sequences of grouped tones. Their findings suggested that both phase locked N2 and sequence-evoked sustained potential are associated with impaired auditory object formation and segmentation in schizophrenia. The result shed light on that the disorder which disrupts normal processing and integration of auditory information in schizophrenia might be associated with the underlying neural processes (Coffman et al., 2016, 2018).

From the perspective of predictive coding theory, the manifestation of N2 was also explained by expectation suppression, which carries the information of prediction error (Han et al., 2019). Alternatively, the sequence evoked sustained response may reflects the encoded precision of inferred sequence predictability (Zhao et al., 2024). Those slow dynamics were hypothesised to be linked to a tonic inhibitory drive that applies gain control to prediction error units, thus dampening responses to predictable stimuli. As a result, the prediction errors are weighted based on their precision, which is manifested in the attenuated N2 response. This interpretation was further supported by emerging evidence suggesting that precision encoding in auditory processing is grounded by an inhibitory mechanism (Natan et al., 2015, 2017; Schulz et al., 2021; Richter and Gjorgjieva, 2022; Yarden et al., 2022).

Except for N2, the findings in this study aligned with the Study 2 which demonstrated that N1 responses (see Figure 5.4A) are modulated by sequence predictability (Hu et al., 2024). However, in both two studies, no correlation between N1 and sustained responses were statistically significant, a confirmation that those two neural mechanisms are more likely to function independently.

How can one reconcile the differences that were observed between N1 and N2? One candidate hypothesis is that the manifestations of prediction errors occur on different layers or time scales, as suggested by models proposing that the cortex generates predictions at various hierarchical levels (Friston, 2005; Kiebel et al., 2008; Garrido et al., 2009; Wacongne et al., 2012). Specifically, the neural processes underlying N1 time interval might encode prediction errors related to spectral temporal details of the sensory attributes (Hosoya et al., 2005; Wacongne et al., 2012). Or the N1 might also encode the low-level statistics which operate on short integration window such as local transition probabilities (Todorovic and Lange, 2012; Maheu et al., 2019). In contrast, the processes underlying N2 might encode errors associated with more complex aspects of the auditory sequence that operate on an integration window across larger time scales or higher layers. This includes the precision of the auditory sequence, as quantified by REG and RND in this study. However, since this study is limited by only two conditions and the local statistics were not rigorously controlled, future research could more systematically manipulate those statistics and use decoding techniques to test this hypothesis.

5.4.4. Temporal Adaptation in N1

Unlike findings from Study 2 that showed statistically reduced N1 amplitudes only in REG condition (Hu et al., 2024), this study reveals that N1 amplitudes decrease over cycles in both the RND and REG conditions. The key to explain this discrepancy may lie in

the encoding of the temporal regularities within the RND and REG sequences. In specific, despite the spectral unpredictability of the RND sequences, the RND sequences may still engage neural mechanisms sensitive to temporal patterns (Costa-Faidella et al., 2011; Hofmann-Shen et al., 2020; van Ackooij et al., 2022) because the timing of sensory input is an important cue for adaptive behaviour. Furthermore, the lack of N1 reduction of RND in Study 2 could be attributed to differences in experimental design, particularly the stimulus presentation rate. One explanation is that the slower rate used in this study has allowed the auditory system more time to encode and integrate temporal regularities, thereby influencing N1 responses. Alternatively, from the perspective of predictive coding, since monitoring slow sequences is computationally demanding, the brain may try to suppress these neural processes once some information within the sensory inputs become predictable (i.e. temporal pattern in this case). This conservation of resources enables it to cope with future uncertainties.

It is important to emphasise that no significant differences in averaged amplitude of N1 activity between REG and RND was seen in Cycle 2 and Cycle 3 after corrected by Cycle 1, while tested by repeated measures ANOVA. This finding is inconsistent with Study 2 (Hu et al., 2024) which demonstrated that the predictability modulates averaged N1 activity. However, a detailed analysis of time-domain responses revealed a significantly reduced response in the N1 time window, as assessed by bootstrap resampling (Figure 5.4C). This was also true for the N2 component, which showed a significantly reduced response in the REG condition that sustained over a certain period but was not tested as significant when averaged the N2 activity across time.

These inconsistencies imply that the observed effects in this study are subtle because the stimuli used are slower and more challenging to track than those in Study 2. Additionally, the participants who effectively responded to the REG pattern were limited in sample size, resulting in a lower signal-to-noise ratio compared to Study 2. Moreover, averaging over time may not fully capture the temporal dynamics of neural processing, especially when the signal-to-noise ratio is low.

5.5. Conclusion

In summary, this study investigated the brain's response to predictable and unpredictable sound sequences in passive listening mode, and the result showed a significant connection between implicit auditory processing and short-term memory. Performance on the Tone Pattern Comparison Task (TP-COMP) correlated with the sustained response to the sound sequence and tone-evoked responses, notably in the N2 time window. This suggests that individuals with better short-term memory can more

effectively distinguish between patterns and random sequences. It implies that the neural circuits for sound pattern detection and short-term memory may intersect, with the hippocampus being a hypothetical region of overlap. Future research could expand on these findings by applying invasive tools with higher spatial resolution to understand the neural mechanisms further. Decoding techniques can also be employed to elucidate the statistical properties represented by those neural activity.

6. General Discussion

6.1. Summary of Main Findings

This PhD thesis investigated the neural dynamics underlying auditory pattern detection process, emphasising the intricate interplay between memory, perception, and sensory processing. The first study explored the influence of informational complexity and temporal duration on auditory memory's role in pattern detection, revealing a significant impact of informational complexity in this cognitive process. Moreover, the behavioural results imply that the brain uses different strategies to integrate sensory inputs with fast and slow presentation rate, enabling efficient comprehension of the auditory environment.

In the second study, MEG was used to measure the brain responses to sound patterns of varying duration. The results suggested that both rapid and slow auditory sequences led to a significant increase in sustained responses to patterned sequences over random ones. Interestingly, single tones in random sequences evoked stronger responses than those in regular sequences. This highlighted the concurrent but opposing effects on the sustained and evoked responses, which jointly shape the neural representation of auditory pattern.

The third study assessed the cognitive factors such as the sustained attention, short-term memory, frequency discrimination, and task engagement to determine their predictability on individual variability in auditory pattern detection. The results consistently indicated that short-term memory capabilities, assessed by the TP-COMP task, significantly predict detection performance across various pattern durations. This reaffirmed the involvement of short-term memory in this cognitive task.

The final study extends this exploration to individuals' brain responses to slow auditory patterns of 5500ms in passive listening mode. The EEG was employed to record the neural responses to patterns, with a particular interest in the correlation between the responses and performance in the TP-COMP task. The study found that individuals with scores above the median in the TP-COMP task showed significantly enhanced neural responses to auditory patterns, compared to those with scores below the median. These findings indicate a shared neural architecture that underpins both auditory short-term memory and auditory scene analysis, providing novel insights into the interconnectedness of perception and cognition. Overall, this series of studies aligns with predictive coding theory

and adds new evidence into the neural mechanism involved in complex auditory pattern recognition, thereby setting the stage for future research.

6.2. Implications for the Brain Functions

This thesis emphasises memory as a crucial function in supporting the auditory system to accumulate sensory evidence. Notably, the memory processes enables the brain to integrate sensory information, and are dynamically modulated based on the duration and timing of these inputs. This adaptability corroborates earlier theoretical models suggesting that the brain operates with intrinsic neural timescales. Such timescales allow various brain regions to process information according to its temporal characteristics, thereby facilitating both the swift assimilation of new sensory data and the extended integration of complex inputs (Hasson et al., 2008; Kiebel et al., 2008; Cavanagh et al., 2020; Golesorkhi et al., 2021). Furthermore, the observations from this thesis are consistent with the hierarchical predictive coding model, which suggests that the brain is structured into multiple layers of processing (Friston, 2008). This structured approach ensures a comprehensive and efficient processing system, optimising the brain's ability to interpret and respond to the changing environment.

The findings suggested that the sustained neural response may serves as the representation of the generative model during the integration process. Previous research hypothesised that these neural correlates were likely to be specific to rapid and continuous sound patterns (Barascud et al., 2016; Southwell et al., 2017; Zhao et al., 2024); however, this thesis provided the novel insights that the neural process also generalises to the analysis of slow and intermittent sound sequences, even when silent gaps are introduced between tones. This is a observation of significance, as silent gaps are typically cues for the auditory system to group sounds and segment events. Rather than merely reflecting the static source of sound or an auditory event, these findings suggested that the sustained response also characterises the dynamic actions or temporal evolution inferred from the sound source. Essentially, the evidence indicated that the sustained response is a common neural mechanism employed by the brain to dynamically generate an internal model, which mirrors the environmental soundscape. From the perspective of predictive coding theory, these neural correlates align with the concept of precision, indicating the inferred reliability of sensory inputs. Psychologically, this is in line with views from a recent review. The review suggested that the sustained neural activity evoked by auditory pattern not only represents the inferred source of the sound but also the actions associated with it, including the temporal predictions (Winkler and Denham, 2024).

Interestingly, the observed correlation between the sustained neural response and short-term memory capabilities suggests that these processes may utilise shared resources or neural pathways. Literatures suggested that the overlapped areas might be sourced to the hippocampus, which are actively involved during both auditory working memory task (Kumar et al., 2016) and implicit sound pattern detection task (Barascud et al., 2016). Latest research by Bonetti and colleagues provide evidence for this hypothesis. In their experiment, participants first familiarised themselves with a musical piece. They were then presented with the original sequences from this piece, as well as the modified versions from the original music pieces. These variations in original piece were introduced at different positions within the sequences, designed to test the participants' ability to detect these deviations. The MEG data identified different neural patterns for processing familiar and novel musical sequences. Multivariate pattern analysis showed that certain brain regions, particularly the superior temporal gyrus and frontal areas, had different activity levels depending on whether the sequences were familiar or new. When the expected musical patterns were disrupted, the auditory cortex quickly generated strong prediction error signals. Furthermore, the hippocampus was more active during the recognition of familiar sequences, highlighting its role in connecting stored auditory memories with sensory input for effective prediction and recognition. The activity in the frontal cortex increased when processing familiar sequences. Error signals and adaptations were also observed in the wider frontoparietal networks, which is consistent with previous literature (Bonetti et al., 2024). These findings, which coincide with the observations in this thesis, highlighted the complex interaction between memory, perception, and cognitive processing within the auditory system. Moreover, those processes are likely to interact dynamically and adaptively, rather than functioning in parallel or within distinct temporal phases.

Those observations are well-supported by the inherently dynamic nature of the auditory environments, which fluctuate according to context and sound sources. Thus, an adaptable and flexible processing system is necessary for efficiently responding to new or changing auditory conditions.

6.3. Limitations and Future Directions

6.3.1. Limitations

The techniques used in this study, including the EEG and the MEG, are known for their excellent temporal resolution. However, their spatial resolution limitations can make it difficult to precisely locate brain activity, particularly in deeper structures like the hippocampus which cannot be fully measured. Their spatial resolution limitations introduces uncertainty on the exact brain regions involved in the observed phenomena.

Also, the inherent noise associated with these methods presents a challenge in signal processing. It is unlikely to completely remove this noise without potentially erasing pertinent signals, and the de-noising techniques (i.e. the DSS method used in this thesis) might unintentionally suppress the signal of interest that were not phase locked to the presentation of stimulus. Moreover, relying on these specific neuroimaging methods may restrict the range of detectable brain activity, possibly missing other relevant neural processes that occur simultaneously but are not captured by EEG or MEG. This could result in an incomplete understanding of the neural dynamics involved.

Furthermore, the stimuli used in this study were limited to pure tone pips, which may not accurately represent the complexity of real-world soundscapes, as these stimuli typically do not consist of simple tones. The stimuli used were chosen for its simplicity, but this choice raises concerns about the difficulty of generalising the conclusions to listening conditions in natural settings. Therefore, it is essential to utilise stimuli that, while still controlled, more closely mimic the spectral and temporal characteristics of natural sounds. This approach will help validate the study's conclusions under more ecologically valid conditions and enhance the generalizability of the results.

6.3.2. Future Direction

The findings of this thesis coincide with the predictive coding framework, supporting the coexistence of multiplexed neural representations in analysing the sensory inputs. The sustained response elicited by the sound sequence correlates with encoding either the predictability of the signal or its precision, while the response evoked by individual tones seems to be tied to prediction errors. Although previous study suggests that the dynamics of the sustained response reflect the concept of precision as suggested by the Bayesian predictive model (Zhao et al., 2024), a quantitative confirmation of this association is still pending. Moreover, it is unclear how sensory prediction and precision interact to underpin the sustained response elicited by the sound sequence, nor how each of these processes contributes to the dynamics observed. The challenge arises as the increased predictability is typically associated with enhanced precision, making it difficult to dissociate. Experimental paradigms could be refined to introduce variations in sensory prediction—targeting intermediate predictability levels—without altering sequence precision or vice versa. This approach would facilitate the precise localisation of brain structures and networks that exhibit differential activation under these specific conditions. For example, prediction representations are more likely to originate from deeper layers, such as the hippocampus, where predictions are presumed to be generated by memory. While the precision is likely to be projected from hippocampus to cortex (i.e. frontal cortex and auditory cortex) and modulate on the post-synapses of prediction error units. Further, employing a regression

model with quantitatively defined precision and prediction parameters as regressors would allow to accurately predict the contribution of each element in sustained responses measured by M/EEG. Alternatively, neural responses often exhibit hierarchical and non-linear dynamics, and hence approaches such as convolutional neural networks might offer more insights for understanding those dynamic interplay between the processes underlying sustained response. Decoding techniques such as the multivariate pattern analysis can be employed to trace sensory prediction across temporal and spatial dimensions. By modelling various layers of prediction—such as transition probabilities, sequence entropy, or contextual statistics—it is possible to identify the specific types of information that the brain processes at distinct regional or temporal points.

Another direction for future research can be placed on understanding the interplay between the tone-evoked response and sustained responses. Evidence presented in this thesis suggests that the N2/N300 component of tone-evoked responses (peak at 300 ms relate to tone onset) are inversely correlated with the amplitude of sustained responses across subjects. Although this correlation appears to support the hypothesis that precision influences prediction error, the two processes are not yet identified as a causal relationship. In addition, these observations suggested that the sustained responses are likely to be related to inhibitory mechanism, though current evidence from this thesis alone is not sufficient to draw definitive conclusions. To investigate and clarify these relationships, future research could adapt the experimental paradigm from Study 4 by shortening the pattern duration (i.e. 5 tones). This modification would ensure that all participants have sufficient memory capacity to effectively track the pattern. Analysis could introduce regression model, using the sustained response as the predictor and the tone response as the dependent variable. This approach can possibly facilitate a more precise analysis of the causal dynamics between these neural activities. Meanwhile, hypotheses related to the types of informational representation underlying N2 and the sustained response can be modelled and tested using the decoding techniques. Dynamic causal modelling could be used to identify the network connection and specific neurons that are related to those two processes.

7. Reference

Aitchison L, Lengyel M (2017) With or without you: predictive coding and Bayesian inference in the brain. Curr Opin Neurobiol 46:219–227.

Albouy P, Mattout J, Bouet R, Maby E, Sanchez G, Aguera P-E, Daligault S, Delpuech C, Bertrand O, Caclin A, Tillmann B (2013) Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. Brain 136:1639–1661.

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

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

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

Ashburner J, Friston KJ (2005) Unified segmentation. NeuroImage 26:839–851.

Auksztulewicz R, Barascud N, Cooray G, Nobre AC, Chait M, Friston K (2017) The Cumulative Effects of Predictability on Synaptic Gain in the Auditory Processing Stream. J Neurosci 37:6751–6760.

Axmacher N, Mormann F, Fernández G, Cohen MX, Elger CE, Fell J (2007) Sustained Neural Activity Patterns during Working Memory in the Human Medial Temporal Lobe. J Neurosci 27:7807–7816.

Baddeley AD, Hitch G (1974) Working Memory. In: Psychology of Learning and Motivation (Bower GH, ed), pp 47–89. Academic Press. Available at: https://www.sciencedirect.com/science/article/pii/S0079742108604521 [Accessed May 9, 2024].

Badre D, D'Esposito M (2007) Functional Magnetic Resonance Imaging Evidence for a Hierarchical Organization of the Prefrontal Cortex. J Cogn Neurosci 19:2082–2099.

Baker CA, Clemens J, Murthy M (2019) Acoustic Pattern Recognition and Courtship Songs: Insights from Insects. Annu Rev Neurosci 42:129–147.

Baldeweg T (2006) Repetition effects to sounds: evidence for predictive coding in the auditory system. Trends Cogn Sci 10:93–94.

Baldeweg T, Richardson A, Watkins S, Foale C, Gruzelier J (1999) Impaired auditory frequency discrimination in dyslexia detected with mismatch evoked potentials. Ann Neurol 45:495–503.

Banai K, Ahissar M (2004) Poor Frequency Discrimination Probes Dyslexics with Particularly Impaired Working Memory. Audiol Neurotol 9:328–340.

Barascud N, Pearce MT, Griffiths TD, Friston KJ, Chait M (2016) Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. Proc Natl Acad Sci 113:E616–E625.

Barbosa LS, Kouider S (2018) Prior Expectation Modulates Repetition Suppression without Perceptual Awareness. Sci Rep 8:5055.

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

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

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw 67:1–48.

Battaglia PW, Jacobs RA, Aslin RN (2003) Bayesian integration of visual and auditory signals for spatial localization. J Opt Soc Am A Opt Image Sci Vis 20:1391–1397.

Baumgarten TJ, Maniscalco B, Lee JL, Flounders MW, Abry P, He BJ (2021) Neural integration underlying naturalistic prediction flexibly adapts to varying sensory input rate. Nat Commun 12:2643.

Bendixen A (2014) Predictability effects in auditory scene analysis: a review. Front Neurosci 8 Available at: https://www.frontiersin.org/articles/10.3389/fnins.2014.00060 [Accessed October 2, 2023].

Benjamin L, Sablé-Meyer M, Fló A, Dehaene-Lambertz G, Roumi FA (2024) Long-Horizon Associative Learning Explains Human Sensitivity to Statistical and Network Structures in Auditory Sequences. J Neurosci 44 Available at: https://www.jneurosci.org/content/44/14/e1369232024 [Accessed June 5, 2024].

Berkes P, Orbán G, Lengyel M, Fiser J (2011) Spontaneous Cortical Activity Reveals Hallmarks of an Optimal Internal Model of the Environment. Science 331:83–87.

Besson M, Schön D, Moreno S, Santos A, Magne C (2007) Influence of musical expertise and musical training on pitch processing in music and language. Restor Neurol Neurosci 25:399–410.

Bianco R, Chait M (2023) No Link Between Speech-in-Noise Perception and Auditory Sensory Memory – Evidence From a Large Cohort of Older and Younger Listeners. Trends Hear 27:23312165231190688.

Bianco R, Hall ETR, Pearce MT, Chait M (2023) Implicit auditory memory in older listeners: From encoding to 6-month retention. Curr Res Neurobiol 5:100115.

Bianco R, Harrison PM, Hu M, Bolger C, Picken S, Pearce MT, Chait M (2020) Long-term implicit memory for sequential auditory patterns in humans Shinn-Cunningham BG, Obleser J, Schroger E, Bieszczad K, eds. eLife 9:e56073.

Bianco R, Mills G, de Kerangal M, Rosen S, Chait M (2021) Reward enhances online participants' engagement with a demanding auditory task. Trends Hear 25:23312165211025941.

Bizley J, Cohen Y (2013) The what, where and how of auditory-object perception. Nat Rev Neurosci 14:693–707.

Blakemore SJ, Wolpert DM, Frith CD (1998) Central cancellation of self-produced tickle sensation. Nat Neurosci 1:635–640.

Bland A, schaefer alexandre (2012) Different Varieties of Uncertainty in Human Decision-Making. Front Neurosci 6 Available at: https://www.frontiersin.org/articles/10.3389/fnins.2012.00085 [Accessed November 30, 2023].

Bonetti L, Fernández-Rubio G, Carlomagno F, Dietz M, Pantazis D, Vuust P, Kringelbach ML (2024) Spatiotemporal brain hierarchies of auditory memory recognition and predictive coding. Nat Commun 15:4313.

Borderie A, Caclin A, Lachaux J-P, Perrone-Bertollotti M, Hoyer RS, Kahane P, Catenoix H, Tillmann B, Albouy P (2024) Cross-frequency coupling in cortico-hippocampal networks supports the maintenance of sequential auditory information in short-term memory. PLOS Biol 22:e3002512.

Bregman AS (1990) Auditory Scene Analysis: The Perceptual Organization of Sound. The MIT Press. Available at: https://direct.mit.edu/books/book/3887/Auditory-Scene-AnalysisThe-Perceptual-Organization [Accessed May 16, 2024].

Buchsbaum BR, Olsen RK, Koch P, Berman KF (2005) Human Dorsal and Ventral Auditory Streams Subserve Rehearsal-Based and Echoic Processes during Verbal Working Memory. Neuron 48:687–697.

Carriere JSA, Cheyne JA, Solman GJF, Smilek D (2010) Age trends for failures of sustained attention. Psychol Aging 25:569–574.

Catchpole CK, Slater PJB (2003) Bird Song: Biological Themes and Variations. Cambridge University Press.

Caucheteux C, Gramfort A, King J-R (2023) Evidence of a predictive coding hierarchy in the human brain listening to speech. Nat Hum Behav:1–12.

Cavanagh SE, Hunt LT, Kennerley SW (2020) A Diversity of Intrinsic Timescales Underlie Neural Computations. Front Neural Circuits 14 Available at: https://www.frontiersin.org/articles/10.3389/fncir.2020.615626 [Accessed January 15, 2024].

Chan RCK, Wang Y, Cheung EFC, Cui J, Deng Y, Yuan Y, Ma Z, Yu X, Li Z, Gong Q (2009) Sustained Attention Deficit Along the Psychosis Proneness Continuum: A Study on the Sustained Attention to Response Task (SART). Cogn Behav Neurol 22:180.

Chandrasekaran B, Kraus N (2010) Music, Noise-Exclusion, and Learning. Music Percept Interdiscip J 27:297–306.

Cheyne JA, Carriere JSA, Smilek D (2009) Absent minds and absent agents: Attention-lapse induced alienation of agency. Conscious Cogn 18:481–493.

Coffman BA, Haigh SM, Murphy TK, Leiter-Mcbeth J, Salisbury DF (2018) Reduced Auditory Segmentation Potentials in First-Episode Schizophrenia. Schizophr Res 195:421–427.

Coffman BA, Haigh SM, Murphy TK, Salisbury DF (2016) Event-related potentials demonstrate deficits in acoustic segmentation in schizophrenia. Schizophr Res 173:109–115.

Costa-Faidella J, Baldeweg T, Grimm S, Escera C (2011) Interactions between "What" and "When" in the Auditory System: Temporal Predictability Enhances Repetition Suppression. J Neurosci 31:18590–18597.

Cowan N (1984) On short and long auditory stores. Psychol Bull 96:341–370.

Cowan N (2008) Chapter 20 What are the differences between long-term, short-term, and working memory? In: Progress in Brain Research (Sossin WS, Lacaille J-C, Castellucci VF, Belleville S, eds), pp 323–338 Essence of Memory. Elsevier. Available at: https://www.sciencedirect.com/science/article/pii/S0079612307000209 [Accessed January 10, 2024].

Cowan N (2019) Short-term memory based on activated long-term memory: A review in response to Norris (2017). Psychol Bull 145:822–847.

Cowan N, Saults JS, Nugent LD (1997) The role of absolute and relative amounts of time in forgetting within immediate memory: The case of tone-pitch comparisons. Psychon Bull Rev 4:393–397.

Crowley KE, Colrain IM (2004) A review of the evidence for P2 being an independent component process: age, sleep and modality. Clin Neurophysiol 115:732–744.

Curtis CE, D'Esposito M (2003) Persistent activity in the prefrontal cortex during working memory. Trends Cogn Sci 7:415–423.

Dalley JW, McGaughy J, O'Connell MT, Cardinal RN, Levita L, Robbins TW (2001) Distinct Changes in Cortical Acetylcholine and Noradrenaline Efflux during Contingent and Noncontingent Performance of a Visual Attentional Task. J Neurosci 21:4908–4914.

Daneman M, Carpenter PA (1980) Individual differences in working memory and reading. J Verbal Learn Verbal Behav 19:450–466.

Dang JS, Figueroa IJ, Helton WS (2018) You are measuring the decision to be fast, not inattention: the Sustained Attention to Response Task does not measure sustained attention. Exp Brain Res 236:2255–2262.

Daunizeau J, Ouden HEM den, Pessiglione M, Kiebel SJ, Stephan KE, Friston KJ (2010) Observing the Observer (I): Meta-Bayesian Models of Learning and Decision-Making. PLOS ONE 5:e15554.

de Cheveigné A, Arzounian D (2018) Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data. NeuroImage 172:903–912.

de Cheveigné A, Parra LC (2014) Joint decorrelation, a versatile tool for multichannel data analysis. NeuroImage 98:487–505.

de Cheveigné A, Simon JZ (2008) Denoising based on spatial filtering. J Neurosci Methods 171:331–339.

de Lange FP, Heilbron M, Kok P (2018) How Do Expectations Shape Perception? Trends Cogn Sci 22:764–779.

Den Ouden HE, Kok P, De Lange FP (2012) How Prediction Errors Shape Perception, Attention, and Motivation. Front Psychol 3 Available at: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2012.00548/full [Accessed June 18, 2024].

Di Liberto GM, Pelofi C, Bianco R, Patel P, Mehta AD, Herrero JL, de Cheveigné A, Shamma S, Mesgarani N (2020) Cortical encoding of melodic expectations in human temporal cortex Peelle JE, Shinn-Cunningham BG, eds. eLife 9:e51784.

Diehl RL, Lotto AJ, Holt LL (2004) Speech Perception. Annu Rev Psychol 55:149–179.

Di Liberto GM, O'Sullivan JA, Lalor EC (2015) Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Curr Biol 25:2457–2465.

Dimakopoulos V, Mégevand P, Stieglitz LH, Imbach L, Sarnthein J (2022) Information flows from hippocampus to auditory cortex during replay of verbal working memory items Griffiths TD, Colgin LL, eds. eLife 11:e78677.

Ding N, Melloni L, Tian X, Poeppel D (2017) Rule-based and word-level statistics-based processing of language: insights from neuroscience. Lang Cogn Neurosci 32:570–575.

Doeller CF, Opitz B, Mecklinger A, Krick C, Reith W, Schröger E (2003) Prefrontal cortex involvement in preattentive auditory deviance detection:: neuroimaging and electrophysiological evidence. NeuroImage 20:1270–1282.

Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. NeuroImage 85:761–768.

Doya K (2007) Bayesian Brain: Probabilistic Approaches to Neural Coding. MIT Press.

Dürschmid S, Edwards E, Reichert C, Dewar C, Hinrichs H, Heinze H-J, Kirsch HE, Dalal SS, Deouell LY, Knight RT (2016) Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. Proc Natl Acad Sci 113:6755–6760.

Dyson BJ, Ishfaq F (2008) Auditory memory can be object based. Psychon Bull Rev 15:409–412.

Efron B, Tibshirani R (1998) An introduction to the bootstrap, Nachdr. Boca Raton, Fla.: Chapman & Hall.

Egermann H, Pearce MT, Wiggins GA, McAdams S (2013) Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. Cogn Affect Behav Neurosci 13:533–553.

Eimas PD, Siqueland ER, Jusczyk P, Vigorito J (1971) Speech Perception in Infants. Science 171:303–306.

Escera C, Alho K, Schröger E, Winkler I (2000) Involuntary attention and distractibility as evaluated with event-related brain potentials. Audiol Neurootol 5:151–166.

Esterman M, Rothlein D (2019) Models of sustained attention. Curr Opin Psychol 29:174–180.

Farrin L, Hull L, Unwin C, Wykes T, David A (2003) Effects of Depressed Mood on Objective and Subjective Measures of Attention. J Neuropsychiatry Clin Neurosci 15:98–104.

Feldman H, Friston K (2010) Attention, Uncertainty, and Free-Energy. Front Hum Neurosci 4 Available at: https://www.frontiersin.org/articles/10.3389/fnhum.2010.00215 [Accessed September 29, 2023].

Ferreira-Santos F (2016) The role of arousal in predictive coding. Behav Brain Sci 39 Available at: https://www.proquest.com/docview/1899322726/abstract/59BD23F5DEFA4C00PQ/1 [Accessed May 30, 2024].

Feuerriegel D, Vogels R, Kovács G (2021) Evaluating the evidence for expectation suppression in the visual system. Neurosci Biobehav Rev 126:368–381.

Fiser J, Aslin RN (2002) Statistical learning of higher-order temporal structure from visual shape sequences. J Exp Psychol Learn Mem Cogn 28:458–467.

Fitzgerald K, Todd J (2018) Hierarchical timescales of statistical learning revealed by mismatch negativity to auditory pattern deviations. Neuropsychologia 120:25–34.

Fletcher H (1940) Auditory Patterns. Rev Mod Phys 12:47–65.

Folstein JR, Van Petten C (2008) Influence of cognitive control and mismatch on the N2 component of the ERP: A review. Psychophysiology 45:152–170.

Fortenbaugh FC, DeGutis J, Esterman M (2017) Recent theoretical, neural, and clinical advances in sustained attention research. Ann N Y Acad Sci 1396:70–91.

Friston K (2002) Functional integration and inference in the brain. Prog Neurobiol 68:113–143.

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

Friston K (2008) Hierarchical Models in the Brain. PLoS Comput Biol 4:e1000211.

Friston K (2009) The free-energy principle: a rough guide to the brain? Trends Cogn Sci 13:293–301.

Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127–138.

Friston K (2012) The history of the future of the Bayesian brain. NeuroImage 62:1230–1233.

Friston K, Chu C, Mourão-Miranda J, Hulme O, Rees G, Penny W, Ashburner J (2008) Bayesian decoding of brain images. NeuroImage 39:181–205.

Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G (2017) Active Inference: A Process Theory. Neural Comput 29:1–49.

Fujioka T, Ross B, Kakigi R, Pantev C, Trainor LJ (2006) One year of musical training affects development of auditory cortical-evoked fields in young children. Brain J Neurol 129:2593–2608.

Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009) The mismatch negativity: A review of underlying mechanisms. Clin Neurophysiol 120:453–463.

Gazzaley A, Cooney JW, Rissman J, D'Esposito M (2005) Top-down suppression deficit underlies working memory impairment in normal aging. Nat Neurosci 8:1298–1300.

Gebhart AL, Newport EL, Aslin RN (2009) Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. Psychon Bull Rev 16:486–490.

Gerhardt HC, Huber F (2002) Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions. Chicago, IL: University of Chicago Press. Available at: https://press.uchicago.edu/ucp/books/book/chicago/A/bo3634687.html [Accessed May 17, 2024].

Giambra LM (1995) A laboratory method for investigating influences on switching attention to task-unrelated imagery and thought. Conscious Cogn 4:1–21.

Giordano BL, McAdams S, Zatorre RJ, Kriegeskorte N, Belin P (2013) Abstract Encoding of Auditory Objects in Cortical Activity Patterns. Cereb Cortex 23:2025–2037.

Glickman M, Usher M (2019) Integration to boundary in decisions between numerical sequences. Cognition 193:104022.

Golesorkhi M, Gomez-Pilar J, Zilio F, Berberian N, Wolff A, Yagoub MCE, Northoff G (2021) The brain and its time: intrinsic neural timescales are key for input processing. Commun Biol 4:1–16.

Gorina-Careta N, Kurkela JLO, Hämäläinen J, Astikainen P, Escera C (2021) Neural generators of the frequency-following response elicited to stimuli of low and high frequency: A magnetoencephalographic (MEG) study. NeuroImage 231:117866.

Graves JE, Pralus A, Fornoni L, Oxenham AJ, Caclin A, Tillmann B (2019) Short- and long-term memory for pitch and non-pitch contours: Insights from congenital amusia. Brain Cogn 136:103614.

Green KP, Kuhl PK, Meltzoff AN, Stevens EB (1991) Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. Percept Psychophys 50:524–536.

Greene CM, Bellgrove MA, Gill M, Robertson IH (2009) Noradrenergic genotype predicts lapses in sustained attention. Neuropsychologia 47:591–594.

Greene RL (1987) Effects of maintenance rehearsal on human memory. Psychol Bull 102:403–413.

Griffiths TD, Hall DA (2012) Mapping Pitch Representation in Neural Ensembles with fMRI. J Neurosci 32:13343–13347.

Griffiths TD, Warren JD (2004) What is an auditory object? Nat Rev Neurosci 5:887–892.

Gussenhoven C (2004) The Phonology of Tone and Intonation. Cambridge University Press.

Han B, Mostert P, de Lange FP (2019) Predictable tones elicit stimulus-specific suppression of evoked activity in auditory cortex. NeuroImage 200:242–249.

Hardt O, Nader K, Nadel L (2013) Decay happens: the role of active forgetting in memory. Trends Cogn Sci 17:111–120.

Harrison PMC, Bianco R, Chait M, Pearce MT (2020) PPM-Decay: A computational model of auditory prediction with memory decay. PLOS Comput Biol 16:e1008304.

Hasson U, Chen J, Honey CJ (2015) Hierarchical process memory: memory as an integral component of information processing. Trends Cogn Sci 19:304–313.

Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N (2008) A Hierarchy of Temporal Receptive Windows in Human Cortex. J Neurosci 28:2539–2550.

Hauser J, Llano López LH, Feldon J, Gargiulo PA, Yee BK (2020) Small lesions of the dorsal or ventral hippocampus subregions are associated with distinct impairments in working memory and reference memory retrieval, and combining them attenuates the acquisition rate of spatial reference memory. Hippocampus 30:938–957.

Hauser MD, Newport EL, Aslin RN (2001) Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. Cognition 78:B53–B64.

Heilbron M, Chait M (2018) Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? Neuroscience 389:54–73.

Helton WS (2009) Impulsive responding and the sustained attention to response task. J Clin Exp Neuropsychol 31:39–47.

Helton WS, Dember WN, Warm JS, Matthews G (1999) Optimism, pessimism, and false failure feedback: Effects on vigilance performance. Curr Psychol 18:311–325.

Helton WS, Hollander TD, Warm JS, Matthews G, Dember WN, Wallaart M, Beauchamp G, Parasuraman R, Hancock PA (2005) Signal regularity and the mindlessness model of vigilance. Br J Psychol Lond Engl 1953 96:249–261.

Helton WS, Kern RP, Walker DR (2009) Conscious thought and the sustained attention to response task. Conscious Cogn 18:600–607.

Helton WS, Weil L, Middlemiss A, Sawers A (2010) Global interference and spatial uncertainty in the Sustained Attention to Response Task (SART). Conscious Cogn 19:77–85.

Henson R, Wakeman D, Litvak V, Friston K (2011) A Parametric Empirical Bayesian Framework for the EEG/MEG Inverse Problem: Generative Models for Multi-Subject and Multi-Modal Integration. Front Hum Neurosci 5:76.

Herrmann B, Buckland C, Johnsrude IS (2019) Neural signatures of temporal regularity processing in sounds differ between younger and older adults. Neurobiol Aging 83:73–85.

Herrmann B, Johnsrude IS (2018) Neural Signatures of the Processing of Temporal Patterns in Sound. J Neurosci 38:5466–5477.

Herrmann B, Maess B, Johnsrude IS (2022) A neural signature of regularity in sound is reduced in older adults. Neurobiol Aging 109:1–10.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Hill PR, Hogben JH, Bishop DMV (2005) Auditory Frequency Discrimination in Children With Specific Language Impairment. J Speech Lang Hear Res 48:1136–1146.

Hofmann-Shen C, Vogel BO, Kaffes M, Rudolph A, Brown EC, Tas C, Brüne M, Neuhaus AH (2020) Mapping adaptation, deviance detection, and prediction error in auditory processing. NeuroImage 207:116432.

Honey CJ, Thesen T, Donner TH, Silbert LJ, Carlson CE, Devinsky O, Doyle WK, Rubin N, Heeger DJ, Hasson U (2012) Slow cortical dynamics and the accumulation of information over long timescales. Neuron 76:423–434.

Hosoya T, Baccus SA, Meister M (2005) Dynamic predictive coding by the retina. Nature 436:71–77.

Houde JF, Jordan MI (2002) Sensorimotor Adaptation of Speech I. J Speech Lang Hear Res 45:295–310.

Hu M, Bianco R, Hidalgo AR, Chait M (2024) Concurrent Encoding of Sequence Predictability and Event-Evoked Prediction Error in Unfolding Auditory Patterns. J Neurosci 44 Available at: https://www.jneurosci.org/content/44/14/e1894232024 [Accessed April 9, 2024].

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

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106-154.2.

Iglesias S, Kasper L, Harrison SJ, Manka R, Mathys C, Stephan KE (2021) Cholinergic and dopaminergic effects on prediction error and uncertainty responses during sensory associative learning. NeuroImage 226:117590.

Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, Ouden HEM den, Stephan KE (2019) Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. Neuron 101:1196–1201.

Ison MJ, Quiroga RQ (2008) Selectivity and invariance for visual object perception. Front Biosci J Virtual Libr 13:4889–4903.

Jasmin K, Lima CF, Scott SK (2019) Understanding rostral–caudal auditory cortex contributions to auditory perception. Nat Rev Neurosci 20:425–434.

Jaunmahomed Z, Chait M (2012) The Timing of Change Detection and Change Perception in Complex Acoustic Scenes. Front Psychol 3 Available at: https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00396 [Accessed June 6, 2023].

Jürgens U (2009) The Neural Control of Vocalization in Mammals: A Review. J Voice 23:1–10.

Justen C, Herbert C (2018) The spatio-temporal dynamics of deviance and target detection in the passive and active auditory oddball paradigm: a sLORETA study. BMC Neurosci 19:25.

Kaiser D, Quek GL, Cichy RM, Peelen MV (2019) Object Vision in a Structured World. Trends Cogn Sci 23:672–685.

Kaiser J (2015) Dynamics of auditory working memory. Front Psychol 6 Available at: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2015.00613/full [Accessed June 6, 2024].

Kaposvari P, Kumar S, Vogels R (2018) Statistical Learning Signals in Macaque Inferior Temporal Cortex. Cereb Cortex 28:250–266.

Karlaftis VM, Giorgio J, Vértes PE, Wang R, Shen Y, Tino P, Welchman AE, Kourtzi Z (2019) Multimodal imaging of brain connectivity reveals predictors of individual decision strategy in statistical learning. Nat Hum Behav 3:297–307.

Kayser C, Petkov CI, Lippert M, Logothetis NK (2005) Mechanisms for allocating auditory attention: an auditory saliency map. Curr Biol CB 15:1943–1947.

Keller TA, Cowan N, Saults JS (1995) Can auditory memory for tone pitch be rehearsed? J Exp Psychol Learn Mem Cogn 21:635–645.

Kersten D, Mamassian P, Yuille A (2004) Object Perception as Bayesian Inference. Annu Rev Psychol 55:271–304.

Kiebel SJ, Daunizeau J, Friston KJ (2008) A Hierarchy of Time-Scales and the Brain. PLOS Comput Biol 4:e1000209.

King AJ, Nelken I (2009) Unraveling the principles of auditory cortical processing: can we learn from the visual system? Nat Neurosci 12:698–701.

Kleiner M, Brainard DH, Pelli D, Ingling A, Murray R, Broussard C (2007) What's new in Psychtoolbox-3. Perception 36:1–16.

Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci 27:712–719.

Knill DC, Saunders JA (2003) Do humans optimally integrate stereo and texture information for judgments of surface slant? Vision Res 43:2539–2558.

Koelsch S, Rohrmeier M, Torrecuso R, Jentschke S (2013) Processing of hierarchical syntactic structure in music. Proc Natl Acad Sci 110:15443–15448.

Koelsch S, Vuust P, Friston K (2019) Predictive Processes and the Peculiar Case of Music. Trends Cogn Sci 23:63–77.

Kok P (2016) Perceptual Inference: A Matter of Predictions and Errors. Curr Biol CB 26:R809-811.

Kok P, de Lange FP (2015) Predictive Coding in Sensory Cortex. In: An Introduction to Model-Based Cognitive Neuroscience (Forstmann BU, Wagenmakers E-J, eds), pp 221–244. New York, NY: Springer. Available at: https://doi.org/10.1007/978-1-4939-2236-9_11 [Accessed June 17, 2024].

Kok P, Jehee JFM, de Lange FP (2012) Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. Neuron 75:265–270.

Kolossa A, Kopp B, Fingscheidt T (2015) A computational analysis of the neural bases of Bayesian inference. NeuroImage 106:222–237.

Kubovy M, Van Valkenburg D (2001) Auditory and visual objects. Cognition 80:97–126.

Kumar M, Federmeier KD, Beck DM (2021a) The N300: An Index for Predictive Coding of Complex Visual Objects and Scenes. Cereb Cortex Commun 2:tgab030.

Kumar S, Bonnici HM, Teki S, Agus TR, Pressnitzer D, Maguire EA, Griffiths TD (2014) Representations of specific acoustic patterns in the auditory cortex and hippocampus. Proc R Soc B Biol Sci 281:20141000.

Kumar S, Gander PE, Berger JI, Billig AJ, Nourski KV, Oya H, Kawasaki H, Howard MA, Griffiths TD (2021b) Oscillatory correlates of auditory working memory examined with human electrocorticography. Neuropsychologia 150:107691.

Kumar S, Joseph S, Gander PE, Barascud N, Halpern AR, Griffiths TD (2016) A Brain System for Auditory Working Memory. J Neurosci 36:4492–4505.

Kumaran D (2008) Short-Term Memory and the Human Hippocampus. J Neurosci 28:3837–3838.

Kumaran D, Maguire EA (2006) An Unexpected Sequence of Events: Mismatch Detection in the Human Hippocampus. PLOS Biol 4:e424.

Lad M, Holmes E, Chu A, Griffiths TD (2020) Speech-in-noise detection is related to auditory working memory precision for frequency. Sci Rep 10:13997.

Lecaignard F, Bertrand O, Caclin A, Mattout J (2022) Neurocomputational Underpinnings of Expected Surprise. J Neurosci 42:474–486.

Lee CS, Aly M, Baldassano C (2021) Anticipation of temporally structured events in the brain Peelen MV, Behrens TE, Geerligs L, eds. eLife 10:e64972.

Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. J Opt Soc Am A Opt Image Sci Vis 20:1434–1448.

Lerner TN, Holloway AL, Seiler JL (2021) Dopamine, Updated: Reward Prediction Error and Beyond. Curr Opin Neurobiol 67:123–130.

Lerner Y, Honey CJ, Katkov M, Hasson U (2014) Temporal scaling of neural responses to compressed and dilated natural speech. J Neurophysiol 111:2433–2444.

Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011) Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. J Neurosci 31:2906–2915.

Li W, Piëch V, Gilbert CD (2004) Perceptual learning and top-down influences in primary visual cortex. Nat Neurosci 7:651–657.

Litvak V, Friston K (2008) Electromagnetic source reconstruction for group studies. NeuroImage 42:1490–1498.

López JD, Litvak V, Espinosa JJ, Friston K, Barnes GR (2014) Algorithmic procedures for Bayesian MEG/EEG source reconstruction in SPM. Neuroimage 84:476–487.

Lü ZL, Williamson SJ, Kaufman L (1992) Human auditory primary and association cortex have differing lifetimes for activation traces. Brain Res 572:236–241.

Luo H, Poeppel D (2012) Cortical Oscillations in Auditory Perception and Speech: Evidence for Two Temporal Windows in Human Auditory Cortex. Front Psychol 3 Available at: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2012.00170/full [Accessed June 26, 2024].

Macleod A, Summerfield Q (1987) Quantifying the contribution of vision to speech perception in noise. Br J Audiol 21:131–141.

Maheu M, Dehaene S, Meyniel F (2019) Brain signatures of a multiscale process of sequence learning in humans de Lange F, Behrens TE, eds. eLife 8:e41541.

Marr D, Vaina L, Brenner S (1997) Representation and recognition of the movements of shapes. Proc R Soc Lond B Biol Sci 214:501–524.

Marshall L, Mathys C, Ruge D, Berker AO de, Dayan P, Stephan KE, Bestmann S (2016) Pharmacological Fingerprints of Contextual Uncertainty. PLOS Biol 14:e1002575.

Mauk MD, Buonomano DV (2004) THE NEURAL BASIS OF TEMPORAL PROCESSING. Annu Rev Neurosci 27:307–340.

May PJC, Tiitinen H (2010) Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Psychophysiology 47:66–122.

Mayrhauser L, Bergmann J, Crone J, Kronbichler M (2014) Neural repetition suppression: evidence for perceptual expectation in object-selective regions. Front Hum Neurosci 8 Available at: https://www.frontiersin.org/articles/10.3389/fnhum.2014.00225 [Accessed May 29, 2024].

McArthur GM, Bishop DVM (2004) Frequency Discrimination Deficits in People With Specific Language Impairment. J Speech Lang Hear Res 47:527–541.

McDermott JH, Schemitsch M, Simoncelli EP (2013) Summary statistics in auditory perception. Nat Neurosci 16:493–498.

McPherson WB, Holcomb PJ (1999) An electrophysiological investigation of semantic priming with pictures of real objects. Psychophysiology 36:53–65.

McWalter R, McDermott JH (2019) Illusory sound texture reveals multi-second statistical completion in auditory scene analysis. Nat Commun 10:5096.

Megela AL, Teyler TJ (1979) Habituation and the human evoked potential. J Comp Physiol Psychol 93:1154–1170.

Mehrpour V, Martinez-Trujillo JC, Treue S (2020) Attention amplifies neural representations of changes in sensory input at the expense of perceptual accuracy. Nat Commun 11:2128.

Menon V, Uddin LQ (2010) Saliency, switching, attention and control: a network model of insula function. Brain Struct Funct 214:655–667.

Menyhart O, Kolodny O, Goldstein MH, DeVoogd TJ, Edelman S (2015) Juvenile zebra finches learn the underlying structural regularities of their fathers' song. Front Psychol 6 Available at: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2015.00571/full [Accessed May 17, 2024].

Meyer T, Olson CR (2011) Statistical learning of visual transitions in monkey inferotemporal cortex. Proc Natl Acad Sci 108:19401–19406.

Miller EK, Cohen JD (2001) An Integrative Theory of Prefrontal Cortex Function. Annu Rev Neurosci 24:167–202.

Miller EK, Erickson CA, Desimone R (1996) Neural Mechanisms of Visual Working Memory in Prefrontal Cortex of the Macaque. J Neurosci 16:5154–5167.

Millman RE, Mattys SL (2017) Auditory Verbal Working Memory as a Predictor of Speech Perception in Modulated Maskers in Listeners With Normal Hearing. J Speech Lang Hear Res JSLHR 60:1236–1245.

Milne A, Bianco R, Poole K, Zhao S, Oxenham A, Billig A, Chait M (2020) An online headphone screening test based on dichotic pitch. Behav Res Methods 53.

Milne A, Zhao S, Tampakaki C, Bury G, Chait M (2021) Sustained pupil responses are modulated by predictability of auditory sequences. J Neurosci Off J Soc Neurosci 41:6116–6127.

Moore BC, Glasberg BR (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J Acoust Soc Am 74:750–753.

Moore BCJ (2014) Development and Current Status of the "Cambridge" Loudness Models. Trends Hear 18:2331216514550620.

Moore BCJ, Gockel HE (2012) Properties of auditory stream formation. Philos Trans R Soc B Biol Sci 367:919–931.

Moore DR, Fuchs PA, Rees A, Palmer AR, Plack C (2010) Oxford Handbook of Auditory Science: Hearing. OUP Oxford.

Mumford D (1994) Neuronal architectures for pattern-theoretic problems. In: Large-scale neuronal theories of the brain, pp 125–152 Computational neuroscience. Cambridge, MA, US: The MIT Press.

Murphy RA, Mondragón E, Murphy VA (2008) Rule Learning by Rats. Science 319:1849–1851.

Näätänen R (1990) The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. Behav Brain Sci 13:201–233.

Näätänen R, Gaillard AWK, Mäntysalo S (1978) Early selective-attention effect on evoked potential reinterpreted. Acta Psychol (Amst) 42:313–329.

Näätänen R, Jacobsen T, Winkler I (2005) Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. Psychophysiology 42:25–32.

Näätänen R, Kujala T, Escera C, Baldeweg T, Kreegipuu K, Carlson S, Ponton C (2012) The mismatch negativity (MMN)--a unique window to disturbed central auditory processing in ageing and different clinical conditions. Clin Neurophysiol Off J Int Fed Clin Neurophysiol 123:424–458.

Näätänen R, Kujala T, Winkler I (2011) Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. Psychophysiology 48:4–22.

Näätänen R, Paavilainen P, Alho K, Reinikainen K, Sams M (1989) Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? Neurosci Lett 98:217–221.

Näätänen R, Picton T (1987) The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. Psychophysiology 24:375–

Nasser HM, Calu DJ, Schoenbaum G, Sharpe MJ (2017) The Dopamine Prediction Error: Contributions to Associative Models of Reward Learning. Front Psychol 8 Available at: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.00244/full [Accessed May 29, 2024].

Natan RG, Briguglio JJ, Mwilambwe-Tshilobo L, Jones SI, Aizenberg M, Goldberg EM, Geffen MN (2015) Complementary control of sensory adaptation by two types of cortical interneurons King AJ, ed. eLife 4:e09868.

Natan RG, Rao W, Geffen MN (2017) Cortical Interneurons Differentially Shape Frequency Tuning following Adaptation. Cell Rep 21:878–890.

Navon D (1977) Forest before trees: The precedence of global features in visual perception. Cognit Psychol 9:353–383.

Nees MA (2016) Have We Forgotten Auditory Sensory Memory? Retention Intervals in Studies of Nonverbal Auditory Working Memory. Front Psychol 7:1892.

Nelken I (2014) Stimulus-specific adaptation and deviance detection in the auditory system: experiments and models. Biol Cybern 108:655–663.

Norman-Haignere SV, Long LK, Devinsky O, Doyle W, Irobunda I, Merricks EM, Feldstein NA, McKhann GM, Schevon CA, Flinker A, Mesgarani N (2022) Multiscale temporal integration organizes hierarchical computation in human auditory cortex. Nat Hum Behav 6:455–469.

Odegaard B, Wozny DR, Shams L (2016) The effects of selective and divided attention on sensory precision and integration. Neurosci Lett 614:24–28.

O'Keeffe F, Dockree P, Moloney P, Carton S, Robertson IH (2007) Awareness of deficits in traumatic brain injury: a multidimensional approach to assessing metacognitive knowledge and online-awareness. J Int Neuropsychol Soc JINS 13:38–49.

Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609.

Olshausen BA, Field DJ (2004) Sparse coding of sensory inputs. Curr Opin Neurobiol 14:481–487.

Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:1:1-1:9.

O'Reilly JX, Jbabdi S, Rushworth MFS, Behrens TEJ (2013) Brain Systems for Probabilistic and Dynamic Prediction: Computational Specificity and Integration. PLOS Biol 11:e1001662.

Ouden HEM den, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010) Striatal Prediction Error Modulates Cortical Coupling. J Neurosci 30:3210–3219.

Owings DH, Morton ES (1998) Animal Vocal Communication: A New Approach. Cambridge: Cambridge University Press. Available at: https://www.cambridge.org/core/books/animal-vocal-communication/D897DF9FEB52AE7C41B6A813B65754AE [Accessed May 17, 2024].

Paavilainen P (2013) The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. Int J Psychophysiol 88:109–123.

Parbery-Clark A, Skoe E, Lam C, Kraus N (2009) Musician enhancement for speech-in-noise. Ear Hear 30:653–661.

Pearce M (2005) The construction and evaluation of statistical models of melodic structure in music perception and composition. City Univ Lond Available at: https://www.academia.edu/1954257/The_construction_and_evaluation_of_statistical_models_of_melodic_structure_in_music_perception_and_composition [Accessed October 5, 2023].

Pearce M, Wiggins G (2004) Improved Methods for Statistical Modelling of Monophonic Music. J New Music Res 33:367–385.

Pearce MT, Ruiz MH, Kapasi S, Wiggins GA, Bhattacharya J (2010) Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. NeuroImage 50:302–313.

Pearce MT, Wiggins GA (2006) Expectation in Melody: The Influence of Context and Learning. Music Percept 23:377–405.

Peebles D, Bothell D (2004) Modelling performance in the Sustained Attention to Response Task. In: Sixth International Conference on Cognitive Modeling. Psychology Press.

Pérez-González D, Lao-Rodríguez AB, Aedo-Sánchez C, Malmierca MS (2024) Acetylcholine modulates the precision of prediction error in the auditory cortex Merchant H, Shinn-Cunningham BG, eds. eLife 12:RP91475.

Pérez-González D, Malmierca M (2014) Adaptation in the auditory system: an overview. Front Integr Neurosci 8 Available at: https://www.frontiersin.org/articles/10.3389/fnint.2014.00019 [Accessed June 6, 2023].

Pessoa L, Kastner S, Ungerleider LG (2003) Neuroimaging Studies of Attention: From Modulation of Sensory Processing to Top-Down Control. J Neurosci 23:3990–3998.

Plack CJ, Moore BC (1990) Temporal window shape as a function of frequency and level. J Acoust Soc Am 87:2178–2187.

Podos J, Huber SK, Taft B (2004) Bird Song: The Interface of Evolution and Mechanism. Annu Rev Ecol Evol Syst 35:55–87.

Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time.' Speech Commun 41:245–255.

Press C, Cook J, Blakemore S-J, Kilner J (2011) Dynamic Modulation of Human Motor Activity When Observing Actions. J Neurosci 31:2792–2800.

Press C, Kok P, Yon D (2020) The Perceptual Prediction Paradox. Trends Cogn Sci 24:13–24.

Quentin R, King J-R, Sallard E, Fishman N, Thompson R, Buch ER, Cohen LG (2019) Differential Brain Mechanisms of Selection and Maintenance of Information during Working Memory. J Neurosci 39:3728–3740.

Ragert M, Fairhurst MT, Keller PE (2014) Segregation and Integration of Auditory Streams when Listening to Multi-Part Music. PLOS ONE 9:e84085.

Randeniya R, Mattingley JB, Garrido MI (2022) Increased context adjustment is associated with auditory sensitivities but not with autistic traits. Autism Res 15:1457–1468.

Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.

Ravignani A, Dalla Bella S, Falk S, Kello CT, Noriega F, Kotz SA (2019) Rhythm in speech and animal vocalizations: a cross-species perspective. Ann N Y Acad Sci 1453:79–98.

Recasens M, Leung S, Grimm S, Nowak R, Escera C (2015) Repetition suppression and repetition enhancement underlie auditory memory-trace formation in the human brain: an MEG study. NeuroImage 108:75–86.

Ren X, Zhang H, Luo H (2022) Dynamic emergence of relational structure network in human brains. Prog Neurobiol 219:102373.

Renoult L, Wang X, Calcagno V, Prévost M, Debruille JB (2012) From N400 to N300: Variations in the timing of semantic processing with repetition. NeuroImage 61:206–215.

Repp BH (1988) Integration and Segregation in Speech Perception. Lang Speech 31:239–271.

Richardson JTE (2007) Measures of Short-Term Memory: A Historical Review. Cortex 43:635–650.

Richter D, Ekman M, de Lange FP (2018) Suppressed Sensory Response to Predictable Object Stimuli throughout the Ventral Visual Stream. J Neurosci 38:7452–7461.

Richter LMA, Gjorgjieva J (2022) A circuit mechanism for independent modulation of excitatory and inhibitory firing rates after sensory deprivation. Proc Natl Acad Sci 119:e2116895119.

Robertson IH, Manly T, Andrade J, Baddeley BT, Yiend J (1997) "Oops!": performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. Neuropsychologia 35:747–758.

Romberg AR, Saffran JR (2010) Statistical learning and language acquisition. Wiley Interdiscip Rev Cogn Sci 1:906–914.

Saffran JR, Aslin RN, Newport EL (1996) Statistical Learning by 8-Month-Old Infants. Science 274:1926–1928.

Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. Cognition 70:27–52.

Sales AC, Friston KJ, Jones MW, Pickering AE, Moran RJ (2019) Locus Coeruleus tracking of prediction errors optimises cognitive flexibility: An Active Inference model. PLoS Comput Biol 15:e1006267.

Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, Formisano E (2014) Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. PLoS Comput Biol 10:e1003412.

Särelä J, Valpola H (2005) Denoising Source Separation. J Mach Learn Res 6:233–272.

Schapiro A, Turk-Browne N (2015) Statistical Learning. In: Brain Mapping, pp 501–506. Elsevier. Available at: https://linkinghub.elsevier.com/retrieve/pii/B9780123970251002761 [Accessed May 16, 2024].

Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, Botvinick MM (2013) Neural representations of events arise from temporal community structure. Nat Neurosci 16:486–492.

Schendan HE, Kutas M (2002) Neurophysiological evidence for two processing times for visual object identification. Neuropsychologia 40:931–945.

Schendan HE, Kutas M (2003) Time course of processes and representations supporting visual object identification and memory. J Cogn Neurosci 15:111–135.

Schendan HE, Kutas M (2007) Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. J Cogn Neurosci 19:734–749.

Schnupp J, Nelken I, King A (2011) Auditory Neuroscience: Making Sense of Sound. MIT Press.

Schnupp JWH, Honey C, Willmore BDB (2013) Neural Correlates of Auditory Object Perception. In: Neural Correlates of Auditory Cognition (Cohen YE, Popper AN, Fay RR, eds), pp 115–149. New York, NY: Springer. Available at: https://doi.org/10.1007/978-1-4614-2350-8_5 [Accessed May 16, 2024].

Schulz A, Miehl C, Berry MJ II, Gjorgjieva J (2021) The generation of cortical novelty responses through inhibitory plasticity Geffen MN, Gold JI, Geffen MN, eds. eLife 10:e65309.

Schulze K, Mueller K, Koelsch S (2011) Neural correlates of strategy use during auditory working memory in musicians and non-musicians. Eur J Neurosci 33:189–196.

Scott SK, Rosen S, Lang H, Wise RJS (2006) Neural correlates of intelligibility in speech investigated with noise vocoded speech—A positron emission tomography study. J Acoust Soc Am 120:1075–1083.

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD (2007) Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. J Neurosci 27:2349–2356.

Seli P, Cheyne JA, Barton KR, Smilek D (2012a) Consistency of sustained attention across modalities: comparing visual and auditory versions of the SART. Can J Exp Psychol Rev Can Psychol Exp 66:44–50.

Seli P, Cheyne JA, Smilek D (2012b) Attention failures versus misplaced diligence: Separating attention lapses from speed–accuracy trade-offs. Conscious Cogn 21:277–291.

Shadmehr R, Krakauer JW (2008) A computational neuroanatomy for motor control. Exp Brain Res 185:359–381.

Shamma S (2001) On the role of space and time in auditory processing. Trends Cogn Sci 5:340–348.

Shamma S (2008) On the Emergence and Awareness of Auditory Objects. PLOS Biol 6:e155.

Shamma SA, Elhilali M, Micheyl C (2011) Temporal coherence and attention in auditory scene analysis. Trends Neurosci 34:114–123.

Shelley EL, Blumstein DT (2005) The evolution of vocal alarm communication in rodents. Behav Ecol 16:169–177.

Shinn-Cunningham BG (2008) Object-based auditory and visual attention. Trends Cogn Sci 12:182–186.

Shipp S (2016) Neural Elements for Predictive Coding. Front Psychol 7 Available at: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2016.01792/full [Accessed May 28, 2024].

Shook A, Marian V, Bartolotti J, Schroeder SR (2013) Musical experience influences statistical learning of a novel language. Am J Psychol 126:95–104.

Skerritt-Davis B, Elhilali M (2018) Detecting change in stochastic sound sequences. PLOS Comput Biol 14:e1006162.

Skerritt-Davis B, Elhilali M (2021) Computational framework for investigating predictive processing in auditory perception. J Neurosci Methods 360:109177.

Skerritt-Davis B, Elhilali M (2021) Neural Encoding of Auditory Statistics. J Neurosci 41:6726–6739.

Smallwood J, Beach E, Schooler JW, Handy TC (2008) Going AWOL in the brain: mind wandering reduces cortical analysis of external events. J Cogn Neurosci 20:458–469.

Smallwood J, Davies JB, Heim D, Finnigan F, Sudberry M, O'Connor R, Obonsawin M (2004) Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. Conscious Cogn 13:657–690.

Smallwood J, McSpadden M, Schooler JW (2007) The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. Psychon Bull Rev 14:527–533.

Snyder JS, Alain C (2007) Toward a neurophysiological theory of auditory stream segregation. Psychol Bull 133:780–799.

Sohoglu E, Chait M (2016) Detecting and representing predictable structure during auditory scene analysis King AJ, ed. eLife 5:e19113.

Solomon SS, Tang H, Sussman E, Kohn A (2021) Limited Evidence for Sensory Prediction Error Responses in Visual Cortex of Macaques and Humans. Cereb Cortex 31:3136–3152.

Southwell R, Baumann A, Gal C, Barascud N, Friston K, Chait M (2017) Is predictability salient? A study of attentional capture by auditory patterns. Philos Trans R Soc B Biol Sci 372:1–26.

Southwell R, Chait M (2018) Enhanced deviant responses in patterned relative to random sound sequences. Cortex J Devoted Study Nerv Syst Behav 109:92–103.

Spierings MJ, Ten Cate C (2016) Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. Proc Natl Acad Sci U S A 113:E3977-3984.

Stefanics G, Kremláček J, Czigler I (2014) Visual mismatch negativity: a predictive coding view. Front Hum Neurosci 8 Available at: https://www.frontiersin.org/articles/10.3389/fnhum.2014.00666 [Accessed May 30, 2024].

Stephens GJ, Honey CJ, Hasson U (2013) A place for time: the spatiotemporal structure of neural dynamics during natural audition. J Neurophysiol 110:2019–2026.

Sterzer P, Adams RA, Fletcher P, Frith C, Lawrie SM, Muckli L, Petrovic P, Uhlhaas P, Voss M, Corlett PR (2018) The Predictive Coding Account of Psychosis. Biol Psychiatry 84:634–643.

Stiso J, Lynn CW, Kahn AE, Rangarajan V, Szymula KP, Archer R, Revell A, Stein JM, Litt B, Davis KA, Lucas TH, Bassett DS (2022) Neurophysiological Evidence for Cognitive Map Formation during Sequence Learning. eneuro 9:ENEURO.0361-21.2022.

Stufflebeam SM, Poeppel D, Rowley HA, L. Roberts TP (1998) Peri-threshold encoding of stimulus frequency and intensity in the M100 latency. NeuroReport 9:91.

Summerfield C, de Lange FP (2014) Expectation in perceptual decision making: neural and computational mechanisms. Nat Rev Neurosci 15:745–756.

Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T (2008) Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci 11:1004–1006.

Sussman ES (2005) Integration and segregation in auditory scene analysis. J Acoust Soc Am 117:1285–1298.

Tabas A, Kriegstein K von (2023) Multiple concurrent predictions inform prediction error in the human auditory pathway. J Neurosci Available at: https://www.jneurosci.org/content/early/2023/11/10/JNEUROSCI.2219-22.2023 [Accessed December 1, 2023].

Takahashi YK, Batchelor HM, Liu B, Khanna A, Morales M, Schoenbaum G (2017) Dopamine Neurons Respond to Errors in the Prediction of Sensory Features of Expected Rewards. Neuron 95:1395-1405.e3.

Takahasi M, Yamada H, Okanoya K (2010) Statistical and Prosodic Cues for Song Segmentation Learning by Bengalese Finches (Lonchura striata var. domestica). Ethology 116:481–489.

Tang MF, Smout CA, Arabzadeh E, Mattingley JB (2018) Prediction error and repetition suppression have distinct effects on neural representations of visual information Summerfield C, Behrens TE, Kok P, Op de Beeck H, eds. eLife 7:e33123.

Temple JG, Warm JS, Dember WN, Jones KS, LaGrange CM, Matthews G (2000) The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. Hum Factors 42:183–194.

Tenenbaum JB, Griffiths TL, Kemp C (2006) Theory-based Bayesian models of inductive learning and reasoning. Trends Cogn Sci 10:309–318.

Thiessen ED (2017) What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. Philos Trans R Soc Lond B Biol Sci 372:20160056.

Thompson RF, Spencer WA (1966) Habituation: a model phenomenon for the study of neuronal substrates of behaviour. Psychol Rev 73:16–43.

Todd J, Michie PT, Schall U, Ward PB, Catts SV (2012) Mismatch negativity (MMN) reduction in schizophrenia—Impaired prediction-error generation, estimation or salience? Int J Psychophysiol 83:222–231.

Todorovic A, Ede F van, Maris E, Lange FP de (2011) Prior Expectation Mediates Neural Adaptation to Repeated Sounds in the Auditory Cortex: An MEG Study. J Neurosci 31:9118–9123.

Todorovic A, Lange FP de (2012) Repetition Suppression and Expectation Suppression Are Dissociable in Time in Early Auditory Evoked Fields. J Neurosci 32:13389–13395.

Tsetsenis T, Broussard JI, Dani JA (2023) Dopaminergic regulation of hippocampal plasticity, learning, and memory. Front Behav Neurosci 16 Available at: https://www.frontiersin.org/articles/10.3389/fnbeh.2022.1092420 [Accessed June 6, 2024].

Turner RS (1977) Hermann von Helmholtz and the empiricist vision. J Hist Behav Sci 13:48–58.

Uddin LQ (2015) Salience processing and insular cortical function and dysfunction. Nat Rev Neurosci 16:55–61.

Uhrig L, Dehaene S, Jarraya B (2014) A Hierarchy of Responses to Auditory Regularities in the Macaque Brain. J Neurosci 34:1127–1132.

Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat Neurosci 6:391–398.

van Ackooij M, Paul JM, van der Zwaag W, van der Stoep N, Harvey BM (2022) Auditory timing-tuned neural responses in the human auditory cortices. NeuroImage 258:119366.

Van de Cruys S, Evers K, Van der Hallen R, Van Eylen L, Boets B, de-Wit L, Wagemans J (2014) Precise minds in uncertain worlds: Predictive coding in autism. Psychol Rev 121:649–675.

Vuust P, Heggli OA, Friston KJ, Kringelbach ML (2022) Music in the brain. Nat Rev Neurosci 23:287–305.

Wacongne C, Changeux J-P, Dehaene S (2012) A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. J Neurosci 32:3665–3678.

Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc Natl Acad Sci 108:20754–20759.

Warm JS, Parasuraman R, Matthews G (2008) Vigilance requires hard mental work and is stressful. Hum Factors 50:433–441.

Watson CS, Foyle DC, Kidd GR (1990) Limits of auditory pattern discrimination for patterns with various durations and numbers of components. J Acoust Soc Am 88:2631–2638.

Weissman DH, Roberts KC, Visscher KM, Woldorff MG (2006) The neural bases of momentary lapses in attention. Nat Neurosci 9:971–978.

Winkler I (2003) Change Detection in Complex Auditory Environment: Beyond the Oddball Paradigm. In: Detection of Change: Event-Related Potential and fMRI Findings (Polich J, ed), pp 61–81. Boston, MA: Springer US. Available at: https://doi.org/10.1007/978-1-4615-0294-4_4 [Accessed May 17, 2024].

Winkler I, Cowan N (2005) From Sensory to Long-Term Memory. Exp Psychol 52:3–20.

Winkler I, Denham SL (2024) The role of auditory source and action representations in segmenting experience into events. Nat Rev Psychol 3:223–241.

Winkler I, Denham SL, Nelken I (2009) Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci 13:532–540.

Winkler I, Karmos G, Näätänen R (1996) Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. Brain Res 742:239–252.

Wood CC, Jennings JR (1976) Speed-accuracy tradeoff functions in choice reaction time: Experimental designs and computational procedures. Percept Psychophys 19:92–102.

Woods DL, Kishiyama MM, Yund EW, Herron TJ, Edwards B, Poliva O, Hink RF, Reed B (2011) Improving digit span assessment of short-term verbal memory. J Clin Exp Neuropsychol 33:101–111.

Wyart V, de Gardelle V, Scholl J, Summerfield C (2012) Rhythmic fluctuations in evidence accumulation during decision making in the human brain. Neuron 76:847–858.

Yabe H, Winkler I, Czigler I, Koyama S, Kakigi R, Sutoh T, Hiruma T, Kaneko S (2001) Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration1. Brain Res 897:222–227.

Yarden TS, Mizrahi A, Nelken I (2022) Context-Dependent Inhibitory Control of Stimulus-Specific Adaptation. J Neurosci 42:4629–4651.

Yon D, Frith CD (2021) Precision and the Bayesian brain. Curr Biol 31:R1026–R1032.

Yu AJ, Dayan P (2005) Uncertainty, Neuromodulation, and Attention. Neuron 46:681–692.

Yu Y, Huber L, Yang J, Fukunaga M, Chai Y, Jangraw DC, Chen G, Handwerker DA, Molfese PJ, Ejima Y, Sadato N, Wu J, Bandettini PA (2022) Layer-specific activation in human primary somatosensory cortex during tactile temporal prediction error processing. NeuroImage 248:118867.

Yuan P (2019) The Neural Code of Working Memory Maintenance. J Neurosci 39:9883–9884.

Zhao S, Skerritt-Davis B, Elhilali M, Dick F, Chait M (2024) Sustained EEG responses to rapidly unfolding stochastic sounds reflect precision tracking. :2024.01.08.574691 Available at: https://www.biorxiv.org/content/10.1101/2024.01.08.574691v2 [Accessed April 22, 2024].

Author Contribution

Chapter 2: The author contributed to the study and experimental stimulus design, data acquisition, data analysis, and writing the manuscript. Maria Chait and Roberta Bianco participated in the study and experimental stimulus design, and give advice in data interpretation.

Chapter 3: The author contributed to the data acquisition, analysis, and writing of the submitted manuscript. Maria Chait participated study and experimental stimulus design, writing of submitted manuscript and provided advice on data analysis. Roberta Bianco participated in data acquisition, offered advice on data analysis, and refined the submitted manuscript. Antonio Rodriguez Hidalgo contributed to the data acquisition and data analysis.

Chapter 4: The author contributed to the study and experimental stimulus design, data acquisition, data analysis, and manuscript writing. Maria Chait participated in study and experimental stimulus design, data analysis and give advice in data interpretation.

Chapter 5: The author was involved in the study and experimental stimulus design, data acquisition and analysis, writing the manuscript. Maria Chait participated in study and experimental stimulus design, and provided advice on data analysis and data interpretation.