Neural Processing of Dynamic Auditory Statistics: How the Passive Listening Brain Responds to Rapidly Changing Sound Environments – Evidence from EEG and Autonomic Responses

Kaho Magami

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 2025

Supervisors:

Professor Maria Chait

Professor Dominik Bach

Declaration

I, Kaho Magami, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Additionally, artificial intelligence has been utilised to correct grammatical errors and refine some sentence structures.

Kaho Magami

July 2025

Abstract

In everyday life, we are immersed in rich and dynamic auditory environments filled with complex statistical regularities. While previous research has demonstrated that the auditory system continuously tracks such patterns, much of this work has relied on simplified stimuli—often consisting of a single, fixed pattern repeated throughout the sequence. In contrast, real-world listening involves constantly shifting auditory patterns embedded in multi-modal sensory contexts.

This PhD thesis investigates how the brain tracks regularities in such environments with dynamic shifts in regularities and explores how these processes influence broader neural and cognitive functions. How do we navigate an uncertain auditory world, and in turn, how does this shape the way we perceive, respond to, and interact with our surroundings? Across three empirical chapters, the work combines electroencephalography (EEG), computational modelling, and psychophysiological measurements to examine different facets of this question. Chapter 2 explores how the brain utilises past information when tracking auditory regularities. By comparing two contexts—one in which past input is informative for predicting the current sequence and one in which it is not—the study investigates whether the brain dynamically adjusts its reliance on memory to form predictions. Chapter 3 investigates whether regularity tracking is influenced by prior context. To test this, the study compares neural responses to identical regular sequences that are preceded by contexts of differing predictability. Chapter 4 examines the broader cognitive impact of automatic auditory regularity tracking. Using an audiovisual memory task, the study tests whether task-irrelevant background sound structures influence the encoding of concurrently presented visual information.

Together, these studies reveal a flexible and context-sensitive mechanism for auditory regularity tracking that operates outside the focus of

attention. This work contributes to our understanding of how the brain maintains adaptive perception in complex, real-world environments.

Impact Statement

This PhD thesis investigates how the human auditory system in healthy young adults processes dynamically evolving sound statistics during passive listening. Our surroundings are filled with complex acoustic patterns that carry rich information about the external world. Yet, scientific understanding of how the brain processes auditory regularities has largely been shaped by studies using static, simplified stimuli that fall short of capturing the adaptive demands of real-world listening. This thesis addresses this gap by examining how the brain tracks evolving statistical regularities across varying contexts, even without directed attention. The findings offer novel insights into the flexible, context-sensitive nature of auditory processing and provide a more ecologically grounded view of how the brain interprets the sounds that surround us. Specifically, the thesis makes three key contributions to the field of cognitive neuroscience:

First, it reveals that the brain's processing of auditory regularities is highly sensitive to the broader environmental context. Rather than relying on a fixed strategy, the brain dynamically adjusts how it uses past information depending on the current context—responding differently to identical sounds based on their surrounding statistical structure. This flexible adjustment in predictive modelling may have direct relevance for understanding psychiatric conditions, such as autism or schizophrenia, where such adaptability to changing environments is often impaired.

Second, the research shows that background regularities are not merely confined to local auditory processing but are integrated into broader cognitive operations. Specifically, the findings suggest that statistical patterns in unattended sounds can influence processes like arousal regulation and memory encoding. This challenges the traditional view of background sounds as irrelevant noise, highlighting their capacity to influence how we perceive and interpret the world, even when our attention is focused elsewhere.

Third, the thesis offers preliminary evidence that the computational principles underlying auditory regularity tracking share similarities with those involved in higher-order cognitive functions such as memory and decision-making. It raises the possibility that the brain's strategy for making sense of background auditory scenes may reflect a more general-purpose mechanism for navigating dynamic, uncertain environments.

Beyond basic science, the work has practical implications for how we design our auditory environments. It highlights the potential cognitive impact of background soundscapes, suggesting that the design of auditory environments—in public spaces, workplaces, educational settings, and digital platforms—should be approached with care. Thoughtfully designed soundscapes could support attention, reduce cognitive load, and enhance wellbeing. By recognising the power of background sound, this work advocates for a more holistic understanding of perception and behaviour—one that accounts for information beyond the focus of attention and reflects the brain's continuous engagement with its sensory environment.

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

1. For a research manuscript that has already been published (if not yet published, please skip to section 2):
   (a) What is the title of the manuscript?
   (b) Please include a link to or doi for the work:
   (c) Where was the work published?
   (d) Who published the work?
   (e) When was the work published?
   (f) List the manuscript's authors in the order they appear on the publication:
   (g) Was the work peer reviewed?
   (h) Have you retained the copyright?
   (i) Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)?
2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3):
   (a) What is the current title of the manuscript?

   'The Effect of Previously Encountered Sensory Information on Neural Representations of Predictability: Evidence from Human EEG'

   (b) Has the manuscript been uploaded to a preprint server e.g. medRxiv'?

   Yes. https://doi.org/10.1101/2025.05.27.656332

   (c) Where is the work intended to be published?

   European Journal of Neuroscience

   (d) List the manuscript's authors in the intended authorship order:

   Kaho Magami, Roberta Bianco, Edward Hall, Marcus Pearce, Maria Chait

(e) Stage of publication:

Under review

3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4):
   KM and MC designed the experiment. KM acquired the data. KM and MC analysed the data. KM, RB, EH, MP, and MC interpreted the results. KM and MC wrote the manuscript. RB, EH, and MP contributed to the revision of the manuscript.
4. In which chapter(s) of your thesis can this material be found?
   Chapter 2

e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate: Kaho Magami
Date: 3.7.2025
Supervisor/Senior Author signature (where appropriate): Maria Chait
Date: 9.7.2025

Acknowledgement

Since stepping into the Ear Institute six years ago as a Master's student, I have had the great fortune to grow—not only as a scientist but also as a person. I am deeply grateful to everyone who has been part of this journey, supporting, challenging, and inspiring me along the way.

First and foremost, I would like to thank Prof. Maria Chait. She has taught me everything about being a researcher—both the hard and soft skills. One of the best and luckiest things in my life has been the opportunity to meet Maria and work closely with her during such a critical period. Her dedication, countless innovative ideas, and passion for science have continually inspired me over the past six years. I am also deeply grateful for her efforts in creating and sustaining an intellectually rich, kind, supportive, and safe research environment.

I also want to express my heartfelt thanks to all past and present members of the Chait Lab for their camaraderie, support, and stimulating conversations. Roberta, Alex, Alice, Sijia, Kat, Adele, and my amazing PhD buddies—Mingyue, Mert, Claudia, Buse, Xena, Hongji, and Xiaoyue—you made this research life not just intellectually rewarding but also filled with laughter, solidarity, and unforgettable memories. I'm truly proud and grateful to have been part of this lab. My thanks also extend to the broader community at the Ear Institute and the EcoBrain team, who have made it a truly special place to learn and grow.

To my collaborators —Prof Marcus Pearce and Prof Toshio Irino— and my thesis committee—Prof Dominik Bach, Prof Lorenzo Picinali, and Prof Gabriella Vigliocco —thank you for your thoughtful guidance, constructive feedback, and steady encouragement throughout this journey. Your insights have challenged and refined my thinking in ways that were truly formative.

Outside the lab, I feel incredibly lucky to have had a loving and supportive “London family.” Thank you, Ayako, Shin, Rie, and the Rose family for being there in every season, and for supporting me through this intense period of life with such warmth and generosity. To all my friends in London—thank you for making this chapter of life vibrant, joyful, and memorable.

Lastly, I want to thank my family and friends in Japan, who have supported me from afar with unwavering love and care. Though distance separated us, your presence has always been with me.

Table of Contents

Declaration **2**

Abstract **3**

Impact Statement **5**

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s) **7**

Acknowledgement **9**

Table of Figures **15**

1. Chapter 1: General Introduction **16**

1.1 Understanding the capability of the passive listening brain 16

1.1.1 Speech 16

1.1.2 Salience detection 21

1.2 Predictive coding theory 25

1.2.1 Introduction to Bayesian brain 25

1.2.2 Predictive coding and its hierarchical structure 27

1.3 Regularity detection, prediction formation, and model update in the auditory system 29

1.3.1 Neural signatures of regularity tracking 30

1.3.2 Computational models 43

1.4 Importance of change point estimation 48

1.4.1 Change point detection in decision making 48

1.4.2 How can the brain achieve change point detection? 49

1.4.3 Unexpected uncertainty as a proxy for change point 51

1.5 Aim of this project 55

2. Chapter 2: The Effect of Previously Encountered Sensory Information on Neural Representations of Predictability: Evidence from Human EEG **60**

2.1 Summary 60
2.2 Introduction 61
2.3 Experiment 1 63
2.3.1 Methods 65
2.3.2 Results and discussion 68
2.4 Experiment 2 73
2.4.1 Methods 77
2.4.2 Results and discussion 80
2.5 General discussion 87
2.5.1 Sustained responses to REG patterns are affected by brief interruptions 87
2.5.2 Sustained response dynamics reflect memory of INT and pre-interruption REG 89
2.5.3 Distinct sustained response patterns in Experiment 1 and Experiment 2 suggest listeners can use or ignore context depending on its relevance 89
3. Chapter 3: The Effect of Prior Context Predictability on the Discovery of New Regularity 95
3.1 Summary 95
3.2 Introduction 95
3.3 Methods 103
3.3.1 Stimuli 103
3.3.2 Procedure 105
3.3.3 Recording and data processing 105
3.3.4 Statistical analysis 108
3.3.5 Modelling 109
3.3.6 Participants 109
3.4 Results 110
3.4.1 The response to the emergence of REGy differs depending on the preceding context 110

3.4.2 The EEG sustained response tracks regularity discovery and violation 113

3.4.3 The prior context influences the discovery process of the following regularity pattern 116

3.4.4 Transition conditions do not yield a block-level effect 119

3.5 Discussion 121

3.5.1 Deviation responses are influenced by the predictability of the prior context 121

3.5.2 REGy discovery dynamics differ between EEG responses and model predictions 125

4. Chapter 4: How Dynamic, Task-Irrelevant Auditory Statistical Changes Shape Visual Memory 130

4.1 Summary 130

4.2 Introduction 131

4.3 Experiment 1 134

4.3.1 Methods 136

4.3.2 Results 144

4.3.3 Discussion 149

4.4 Experiment 2 155

4.4.1 Methods 155

4.4.2 Results 160

4.4.3 Discussion 163

4.5 General discussion 166

4.5.1 The role of surprise and context change in memory boundary formation 167

4.5.2 Broader cognitive effects of background sound 169

5. Chapter 5: General Discussion 172

5.1 Summary of findings 172
5.2 Implications 173
5.3 Limitations 176
5.4 Future direction 178
6. Appendix Chapter: Is Speaker Size a Salient Auditory Feature? 181
6.1 Summary 181
6.2 Introduction 181
6.3 Experiment 1 184
6.3.1 Methods 184
6.3.2 Results 186
6.3.3 Discussion 189
6.4 Experiment 2 190
6.4.1 Methods 193
6.4.2 Results 196
6.4.3 Discussion 199
References 202
Author Contribution 235

Table of Figures

Figure 1.1 MEG responses to rapidly evolving statistical structures in sound stimuli. 39
Figure 1.2 Model illustrations. 45
Figure 1.3 Summary of current understanding of sustained neural responses. 58
Figure 2.1 Experiment 1: stimuli, model simulations, and EEG results. ... 71
Figure 2.2 Experiment 2: stimuli and model simulations. 77
Figure 2.3 Experiment 2: INT evoked deviance response. 81
Figure 2.4 Experiment 2: sustained response dynamics. 84
Figure 2.5 Comparing EEG across Experiments. 91
Figure 2.6 Results from Bianco et al. (2025) reveal results consistent with Experiment 2 here. 93
Figure 3.1 Stimuli and model simulations. 103
Figure 3.2 Transition evoked deviance responses. 112
Figure 3.3 Sustained response dynamics. 115
Figure 3.4 REGy discovery dynamics. 119
Figure 3.5 Comparisons of control conditions across blocks. 120
Figure 4.1 Stimuli and task schematics. 143
Figure 4.2 Behavioural results of the influence of the background sound on the ongoing visual task. 145
Figure 4.3 Skin conductance activity measured during the encoding session. 149
Figure 4.4 Stimuli and task schematics. 159
Figure 4.5 Behavioural results of the influence of the background sound on the visual task. 162
Figure 6.1 Results of the subjective size judgement. 187
Figure 6.2 Judgement matrix for each sound pair. 189
Figure 6.3 Examples of sound-evoked ocular dynamics. 193
Figure 6.4 Ocular dynamics evoked by sounds conveying different speaker size. 198

1. Chapter 1: General Introduction

1.1 Understanding the capability of the passive listening brain

Imagine walking through King's Cross Station: a departure announcement echoes across the concourse, heels tap hurriedly past, a group of tourists chatter excitedly in a mix of accents. Outside, a street performer begins a tune on a saxophone, overshadowed by the blaring siren of an ambulance. Each sound source — transient or sustained, speech or mechanical — emerges, shifts, and fades unpredictably, across time, space, and spectral range.

We are constantly immersed in such complex soundscapes and effortlessly navigate this dynamic flow of information. Yet, our understanding of how such environments influence the brain remains limited. This thesis explores how the brain adapts to, responds to, and learns from these rich and ever-changing auditory signals through exposure. I begin the thesis by reviewing the kinds of information that the passive listening brain can extract and learn from the surrounding sound environment, specifically focusing on two major functions of listening: speech processing and salience detection.

1.1.1 Speech

Voice perception is a fundamental and widespread ability—not only in humans, but in many animal species that rely on complex vocal communication systems (Gil-da-Costa et al., 2004; Harford et al., 2024; Petkov et al., 2008). From early infancy, humans begin to acquire core structural properties of their native language—such as prosodic contours, phonotactic constraints, and word boundaries—through exposure, without explicit instruction (Jusczyk, Cutler, et al., 1993; Jusczyk, Friederici, et al., 1993; Jusczyk et al., 1994; Saffran, Aslin, et al., 1996; Wellmann et al., 2012).

A central mechanism supporting such acquisition is statistical learning — the ability to detect regularities in the sensory environment through passive exposure. In the context of speech, this involves learning patterns embedded in the signal, such as the likelihood of one sound following another or the frequency distributions of certain phonetic features. Crucially, statistical learning is typically implicit: it operates without deliberate instruction or conscious awareness, and learners may not experience a subjective sense of familiarity with the patterns they have acquired (Turk-Browne et al., 2009).

Even after the foundational rules of language are acquired, speech perception remains a dynamic task. Spoken language is inherently variable, influenced by accents, dialects, talker identity, and noisy environments. These variations shift the reliability and relevance of acoustic dimensions, requiring listeners to continuously sample their auditory environment and adjust their internal models to accommodate the present context (Holt, 2025; Holt et al., 2018). The following sections review the range of speech regularities that listeners learn through exposure, and the flexibility of such representations.

1.1.1.1 Transitional probability

A seminal demonstration of statistical learning comes from the work of Saffran, Aslin, et al. (1996), who showed that 8-month-old infants could extract word-like units from a continuous stream of artificial speech by tracking transitional probabilities (TPs)—that is, the probability that one syllable follows another. In their study, infants were exposed to a two-minute stream in which tri-syllabic “words” had high internal TPs (e.g., syllable B always followed syllable A), but low TPs at word boundaries. In a subsequent test phase, infants successfully distinguished the learned “words” from non-words. This finding has been extended to adults (Saffran, Newport, et al., 1996), neonates (Teinonen et al., 2009), and natural languages (Pelucchi et al., 2009), suggesting that humans are equipped with a robust, passive ability to detect TP from infancy.

1.1.1.2 Non-adjacent learning

Natural language is replete with non-adjacent regularities, where relevant elements are separated by intervening material. While more complex than adjacent dependencies, such patterns can also be acquired through exposure. Newport and Aslin (2004) found that adults could learn non-adjacent dependencies, although in this experiment, the task required active engagement with the stimuli. Furthermore, Friederici et al. (2011) showed that 4-month-old infants could learn such patterns passively, suggesting that the brain is tuned to structure beyond local transitions.

1.1.1.3 Distributional learning

Another powerful learning mechanism is distributional learning, which allows listeners to track how often particular features or cues occur in the speech signal. For instance, while young infants can discriminate nearly all phonetic contrasts (Aslin et al., 1998), by the end of their first year their perception becomes attuned to the phonemic structure of their native language—much like adults (Kuhl et al., 1992; Werker & Tees, 1984). This shift is considered to reflect sensitivity to the statistical distribution in their native language (Guenther & Gjaja, 1996; Kohl, 1993). Supporting this, Maye et al. (2002) demonstrated that 6-month-old infants could shift their sound discrimination patterns after brief passive exposure to bimodal or unimodal distributions of phonetic variation.

This learning process also shapes how listeners weight different acoustic dimensions. When categorising speech sounds, the brain must integrate information across multiple acoustic cues—but not all cues are equally informative. Listeners learn, through long-term exposure, to prioritise dimensions that most reliably distinguish categories (Holt & Lotto, 2006; Toscano & McMurray, 2010; Wellmann et al., 2012). For example, when distinguishing /r/ and /l/, American listeners rely heavily on the onset frequency of the third formant (F3), while Japanese listeners— whose native language

does not contrast these sounds— tend to rely more on the second formant (F2), a less reliable cue in this context (Iverson et al., 2003).

1.1.1.4 Co-occurring feature learning

Speech signals are multidimensional, and infants also learn through exposure to co-occurring cues. Results from Thiessen and Saffran (2003) suggest that infants initially rely on transitional probabilities to segment words from continuous speech. Over time, they begin to integrate co-occurring stress cues and gradually shift their weighting toward such cues as they develop.

Crucially, real-world learning often involves ambiguity. For example, when a word is presented, infants may see multiple potential referent objects, creating uncertainty about the correct mapping. Smith and Yu (2008) demonstrated that 12-months-old infants can learn word-object mappings even when multiple objects and words are presented simultaneously. Through repeated exposure across trials, infants were able to statistically infer the correct pairings—highlighting the brain's capacity to extract consistent mappings from noisy environments.

1.1.1.5 Generalisation of learned structure

Statistical learning supports not only the recognition of specific regularities but also the generalisation of learned regularities. Marcus et al. (1999) exposed 7-months-old infants to sequences with structures like ABA (e.g., "ga ti ga") or ABB ("ga ti ti"), then tested them on novel sequences using different syllables. Infants successfully recognised the familiar pattern, demonstrating sensitivity to underlying structure, not just surface features. Further evidence from Gómez and Lakusta (2004) shows that such generalisation can occur even under noisy conditions— when 17% of training instances came from a different language—highlighting the robustness of statistical learning mechanisms.

1.1.1.6 Long-term vs. short-term learning

As reviewed above, long-term exposure to the statistical structure of speech shapes how listeners perceive and categorise sounds. For instance, listeners learn to prioritise acoustic dimensions that are most consistently informative in their long-term exposure through development (Idemaru & Holt, 2013; Iverson et al., 2003; Kondaurova & Francis, 2008).

However, speech signals—and the environments in which they occur—are inherently variable. Rather than relying on fixed representations, the perceptual system is remarkably flexible. A growing body of evidence suggests that listeners can rapidly adjust their perceptual strategies aligning with the short-term input regularities, even if it contradicts with the regularities familiarised through long-term exposure (Holt, 2025; Holt et al., 2018).

In a study by Holt and Lotto (2006), participants were trained to categorise sounds that varied along two acoustic dimensions: carrier frequency (CF) and modulation frequency (MF). Most listeners naturally weighted CF more heavily, as it was more diagnostic for categorisation. However, a separate group that received brief pre-exposure to a distribution in which only MF varied (while CF remained constant) showed an increased reliance on MF. This shift occurred through passive exposure alone, suggesting that listeners implicitly track recent statistical patterns and adjust their weighting of cues accordingly.

Such adjustments are not limited in the lab. In real-world scenarios, such as encountering a foreign accent, long-term perceptual expectations often fail to align with the input. Accented speech can distort familiar cue distributions, requiring rapid recalibration. Research shows that listeners can flexibly modify the weight given to specific acoustic cues when exposed to accented speech, effectively adapting to the altered structure of the signal (Hodson et al., 2023; Idemaru & Holt, 2011; T. K. Murphy et al., 2024).

Together, these findings illustrate that speech perception reflects a dynamic interplay between long-term regularity learning and short-term adaptation. The brain continuously integrates prior experience with current input, enabling flexible and context-sensitive interpretation of speech sound.

1.1.2 Salience detection

Another crucial function of the auditory system is the rapid detection of potential danger. Unlike the visual system, which typically requires directed gaze, the auditory system continuously monitors the environment regardless of our attentional focus. Moreover, auditory stimuli elicit faster responses related to attentional orientation toward potential danger compared to visual stimuli, positioning the auditory system as the brain's early warning mechanism (Dalton & Lavie, 2004; S. Murphy et al., 2013; Rolfs et al., 2008; C.-A. Wang et al., 2014).

A key concept in this context is sound salience—the automatic attraction of attention by sound. Salient sounds can trigger a shift in attention even when processed in the background, often signalling an urgent or potentially threatening event. This ability allows the auditory system to alert us to sudden changes or anomalies in the environment that might require immediate action.

Salience is shaped by both the acoustic properties of the sound and its relation to the surrounding auditory context. For example, a car horn is inherently salient, but it is likely to be perceived as more salient in a quiet countryside than in a bustling city, where such sounds are more expected. In the following sections, I review key factors that shape auditory salience—that is, the features that engage the passive listening brain.

1.1.2.1 Salient sound components

Consider the design of an auditory alarm: these sounds are purposefully made to capture attention. They are often jarring, unpleasant, and difficult to

ignore. Regardless of the surrounding context, such sounds tend to “pop out” perceptually. But what makes a sound so hard to ignore?

Certain acoustic features inherently contribute to sound salience. Through both evolutionary adaptation and learned experience, the brain has developed heightened sensitivity to specific sound properties that signal importance, urgency, or danger.

Traditionally, sound salience has been measured using subjective judgments, where listeners are asked to rate how distracting or attention-grabbing a sound is. However, recent developments in physiological measurement techniques have introduced more objective ways of quantifying salience while participants passively processing the sound. Notably, ocular dynamics—specifically pupil dilation response (PDR) and microsaccadic inhibition (MSI)—have gained traction as reliable indices. PDR is thought to reflect general arousal levels, while MSI, referred to as the “ocular freezing” response, indicates a shift in attentional allocation (Bonneh et al., 2015; Rolfs, 2009; Sara, 2009; Sara & Bouret, 2012). A more detailed review of these physiological measures is provided in the appendix chapter.

Among the acoustic components that influence salience, loudness is perhaps the most intuitive. Loud sounds tend to be more difficult to ignore. Empirical findings support this intuition: For instance, Liao et al. (2016) demonstrated a correlation between loudness and perceived salience. Furthermore, louder sounds were associated with larger PDR (Liao et al., 2016; see also N.Huang and Elhilali (2017)).

Beyond loudness, another important acoustic factor is roughness. Roughness refers to a perceptual quality typically described as harsh, raspy, or buzzing (Zhao, Wai Yum, et al., 2019). Human screams—a natural alarming sound—cluster within the roughness range in acoustic space (Arnal et al., 2015). This quality is associated with energy in the high-frequency range of the

amplitude modulation spectrum, specifically above 30 Hz (Arnal et al., 2015; Zhao, Wai Yum, et al., 2019). Zhao, Wai Yum, et al. (2019) showed that roughness was correlated with subjective salience rating, and that it also correlated with MSI responses. Interestingly, neuroimaging data revealed that rough sounds activated the amygdala, a brain region implicated in emotion and danger processing (Arnal et al., 2015). This suggests the importance of roughness sound component for danger detection.

Another potent cue for salience is the looming sound—sounds that appear to be approaching the listener in space. Looming sounds carry survival-relevant information, such as an approaching predator in nature or an oncoming vehicle in urban environments. Numerous studies have shown that looming sounds are perceived as more salient than the receding sound, the reversed version of the looming sound that represent the objects moving away from the listener. For instance, listeners detect looming sounds more rapidly and consistently rate them as louder, closer, and longer in duration compared to identical sounds played in reverse (Bidelman & Myers, 2020; Grassi & Darwin, 2006; Neuhoff, 2001; Ponsot et al., 2015). Looming stimuli also elicit stronger physiological responses reflecting heightened phasic alertness and engage the amygdala, again linking them to neural circuits responsible for monitoring potential threats (Bach et al., 2008).

In the appendix chapter, I further review and examine additional types of potentially salient features.

1.1.2.2 Salience relative to the surrounding context

While these acoustic features are inherently attention-capturing, sound salience is also strongly modulated by the surrounding auditory context. A sound that is loud or rough may appear less salient if embedded in a background with similar features. In other words, salience is not determined solely by absolute acoustic properties, but also by the extent to which a sound

deviates from the statistical regularities of the ongoing auditory scene (N. Huang & Elhilali, 2017). According to predictive coding theory (Friston, 2010), the brain constructs internal models of the sensory environment by continuously extracting regularities from input. These models enable the detection of unexpected events—those that violate the predictions generated by prior context.

As empirical support for this context-based view of salience, Kaya et al. (2020) constructed melodies in which the acoustic dimensions followed specific statistical distributions. Occasionally, a note would deviate from these distributions, serving as a salient event. Importantly, the same notes were also embedded in alternative melodies where they conformed to the distribution, allowing for a controlled comparison. Neural responses were recorded while participants focused on a separate visual task, ensuring that attention was not deliberately directed toward the sounds. Despite this, deviant notes elicited stronger neural signatures indicating an enhanced allocation of attention to the sound than their acoustically identical counterparts. In a related experiment, Kaya and Elhilali (2014) asked participants to identify salient events within sound clips and compared their responses to predictions from a computational model based on predictive coding. This model estimated the likelihood of future events from prior acoustic regularities and flagged events deviating from such predictions. Strikingly, its predictions closely matched participants' behavioural judgments, further supporting the notion that salience is determined by deviations from expected statistical patterns.

In summary, the brain's ability to extract regularities, generate predictions, and detect deviations lies at the heart of both salience detection and speech perception—even in the absence of focused attention. This dynamic interplay allows listeners to navigate and interpret complex auditory environments with remarkable efficiency. The following sections delve deeper into the mechanisms underlying regularity extraction and predictive processing, which form the foundation of passive auditory perception.

1.2 Predictive coding theory

As reviewed above, perception is not a passive reception of sensory input, but an active process of inference. The brain continuously seeks to uncover the underlying regularities in the environment, allowing it to utilise such regularities to interpret sensory signals. This inferential capacity is crucial because sensory input is often uncertain—any given stimulus could arise from multiple possible causes (Knill & Pouget, 2004). By forming predictions about the world, the brain not only makes sense of ambiguous input but also gains the ability to respond more rapidly and efficiently to external events (Bendixen et al., 2012; Boubenec et al., 2017; Nobre et al., 2007; Southwell & Chait, 2018).

This predictive process is formalised in the predictive coding theory, which proposes that the brain builds internal models of the environment, continuously compares these models against incoming sensory data, and updates them based on the discrepancy—known as *prediction error*—between expectation and reality (Friston, 2005, 2008; Y. Huang & Rao, 2011; Rao & Ballard, 1999). The following section explores how such internal models are generated and refined in the brain.

1.2.1 Introduction to Bayesian brain

A growing body of evidence suggests that predictive coding operates according to the principles of Bayesian inference (Friston, 2005, 2008; Knill & Pouget, 2004; Skerritt-Davis & Elhilali, 2018; but see Aitchison & Lengyel, 2017). The Bayesian framework aligns with the long-standing view, dating back to Helmholtz, that perception is a form of unconscious inference—an interpretative process that relies on prior knowledge to resolve the inherent ambiguity of sensory input (Kersten et al., 2004).

Bayesian inference provides a formal framework for describing how the brain integrates prior knowledge (top-down expectations) with new sensory

evidence (bottom-up input) to update the belief. This process is captured by Bayes' theorem:

$$P(H|E) \propto P(E|H)\,P(H)$$

Here, H represents a **Hypothesis** or prediction, and E is the incoming sensory **Evidence**. The model evaluates how likely the evidence is, given the hypothesis (*P(E/H)* or likelihood), and combines it with the prior probability of that hypothesis (*P(H)*) to generate an updated belief, the *posterior* (*P(H/E)*). The regularities learned through long-term experience, as discussed in the previous section, provide the basis for these prior expectations.

Consider a simple example: you hear barking in your yard (*E*), and you wonder if it is your dog (*H*). To evaluate this hypothesis, you assess how strongly the evidence supports it—that is, you consider the probability of H given E, or *P(H|E)*. This depends on two things: how likely it is that your dog would be barking at all (*P(H)*), and how likely it is that your dog would produce the particular bark you heard (*P(E|H)*). The more the sound matches what you expect from your dog, the more confident you become that your interpretation is correct.

Importantly, Bayesian inference operates on *distributions* (Knill & Pouget, 2004); distribution with larger variance represents the inherent uncertainty of that information. When sensory input is noisy or ambiguous (high variance), its associated likelihood distribution *P(E|H)* has a high variance, and contributes less to the posterior *P(H|E)*. As a result, the prior *P(H)* plays a more dominant role in shaping the updated belief. Conversely, when the sensory signal is precise, it has more power to update existing predictions (Kersten et al., 2004; Yu & Dayan, 2005). This inferential process involves continuously maintaining and evaluating multiple hypotheses, updating the internal model to reflect the most probable interpretation of the environment. The Bayesian framework thus enables to adaptively revise its representations in response to ongoing sensory input—a key advantage for navigating an ever-changing world.

1.2.2 Predictive coding and its hierarchical structure

While the Bayesian framework offers a theoretical account of how the brain could update its beliefs in light of new sensory evidence, predictive coding provides a plausible neural mechanism for implementing this process (Aitchison & Lengyel, 2017; Bastos et al., 2012). In particular, hierarchical predictive coding has gained prominence as a model that explains how the brain organises perception and learning.

Hierarchical predictive coding proposes that perception arises from the continuous interaction between higher-level predictions and lower-level sensory input, through feedforward and feedback connections (Friston, 2005, 2008; T. S. Lee & Mumford, 2003; Mumford, 1992; Rao & Ballard, 1999; Shipp, 2016). Higher cortical areas generate predictions about expected sensory input and send them down, while lower sensory regions compare these predictions to the actual input. Any mismatch—termed a prediction error—is then transmitted upward. For instance, predictions formed in visual area V2 are sent back to V1 via feedback projections and serve as prior expectations in V1. V1 compares this prediction with the actual incoming signal; any unexplained residual—i.e., the prediction error—is sent forward to V2, prompting an update of the internal model (Rao & Ballard, 1999). This iterative loop continues across the cortical hierarchy, with the goal of minimising prediction error and optimising the internal model of the world (Friston, 2005, 2008).

The viability of hierarchical predictive coding is supported by the brain's known anatomical organisation, which exhibits a hierarchical structure across cortical areas (Felleman & Van Essen, 1991; Zeki & Shipp, 1988). Within this hierarchical structure, distinct neural pathways are proposed to mediate the flow of predictions and prediction errors. Feedforward pathways, which carry prediction error signals, typically involve excitatory projections from pyramidal neurons in the superficial layers. In contrast, feedback pathways that transmit top-down predictions typically arise from deep-layer pyramidal neurons and

modulate activity in lower cortical areas, often through inhibitory connections (Bastos et al., 2012; Shipp, 2016).

1.2.2.1 Precision-weighted prediction error

Prediction error lies at the heart of hierarchical predictive coding models, serving as the primary signal that drives the updating of internal representations to align more closely with the environment. In theory, prediction errors are informative: they reflect meaningful discrepancies between top-down expectations and bottom-up sensory input, indicating that the brain's current model of the world may need revision. However, the real world is inherently noisy, and not all prediction errors stem from meaningful discrepancies—some arise purely from sensory noise. This raises a critical question: how does the brain distinguish between informative prediction errors that warrant updating, and spurious ones that should be ignored?

A central solution proposed by predictive coding frameworks is the concept of precision, defined as the inverse of variance. Precision quantifies the estimated reliability of a prediction. By weighting prediction errors according to precision of the prediction, the brain can regulate their influence on model updates (Friston, 2008; Yon & Frith, 2021). As discussed above, prediction error arises from the comparison between sensory input and prediction. Therefore, the reliability of both sources critically determines the weight assigned to the resulting error signal. For example, when sensory input is noisy, it results in a broader predictive distribution, causing the associated prediction error to receive low precision and therefore exert a small influence on updating the internal model. In contrast, when prior beliefs are strong and reliable, the predictive distribution becomes narrower, assigning higher precision weighing to the prediction error and increasing its impact on model updating. In this way, precision functions as a modulatory parameter that dynamically adjusts the influence of sensory evidence versus prior expectations (Friston, 2008; Yon & Frith, 2021). At the neural level, precision is believed to be encoded by

modulating the postsynaptic gain of superficial pyramidal neurons which encode prediction error (Feldman & Friston, 2010; Friston, 2008; Shipp, 2016).

Such mechanism enables the brain to maintain adaptive and efficient internal models in complex, ever-changing environments. Importantly, disruptions in the estimation or use of precision may underlie certain neuropsychological conditions. For example, in autism, atypically strong belief about the precision of incoming sensory evidence may cause the system to overreact to minor fluctuations, prompting excessive model updates (Lawson et al., 2014; Pellicano & Burr, 2012; Yon & Frith, 2021; but see Lieder et al., 2019).

In summary, this section reviewed the core principles of predictive coding theory—a foundational framework for understanding how the brain interprets and adapts to sensory input. In the following section, I turn to the auditory domain, examining the empirical evidence for whether and how predictive coding can be implemented in the auditory system.

1.3 Regularity detection, prediction formation, and model update in the auditory system

As discussed in Chapter 1.1, the ability to detect regularities and form predictions based on learned patterns is a fundamental function of the auditory system—even when sounds are not directly tied to behaviour. Empirically testing this ability in passive listeners, however, presents unique challenges. In decision-making tasks, participants' choices provide a window into their expectations about the environment. In contrast, during passive listening in the auditory domain, it is much harder to determine what regularities have been learned or when a listener experiences a prediction violation. This section reviews the indirect methods that have been developed over the past decades to investigate how the auditory system tracks regularities and processes prediction models.

1.3.1 Neural signatures of regularity tracking

To investigate how the auditory system computes predictions, electroencephalography (EEG) and magnetoencephalography (MEG) are particularly valuable tools, owing to their high temporal resolution. Since auditory scenes unfold rapidly over time, capturing the precise timing of neural responses is essential. Using these techniques, researchers have developed two complementary approaches to uncover the neural correlates of regularity detection and prediction formation: one focuses on identifying neural signatures of predictions themselves, while the other examines responses to deviations from those predictions.

1.3.1.1 Repetition suppression

When auditory stimuli are repeatedly presented, neural responses to those stimuli often diminish—a phenomenon known as repetition suppression (Baldeweg, 2006; Todorovic & de Lange, 2012). A traditional explanation for this effect is neuronal adaptation, where the reduced activity reflects passive fatigue or habituation of sensory neurons. However, growing evidence suggests that repetition suppression cannot be fully explained by such passive mechanisms alone and may also reflect active, expectation-related processes.

For instance, Costa-Faidella et al. (2011) examined repetition suppression in a passive listening context by presenting sequences of pure tones with either a predictable (fixed) or unpredictable (varied) inter-stimulus interval (ISI). Despite equivalent levels of sensory exposure—and thus similar opportunities for adaptation—the unpredictable condition showed reduced repetition suppression. This indicates that predictability itself can modulate neural responses, suggesting that repetition suppression may, in part, reflect the operation of active predictive mechanisms.

A more direct dissociation between adaptation and expectation was demonstrated by Todorovic and de Lange (2012), who orthogonally manipulated adaptation (via tone repetition) and expectation (via transitional probabilities).

For example, repetitions of tone A were highly predictable (A followed by A was common), whereas repetitions of tone B were less predictable (B followed by B was rare), allowing the same physical repetition to carry different levels of expectation. Their findings showed that both repetition and expectation suppressed neural responses, but with distinct temporal profiles: repetition effects (repeated vs alternating tones) appeared earlier in the neural response, while expectation effects (expected vs unexpected tones) emerged later. This result suggests that neural suppression arises not only from adaptation but also from predictive processes.

Moreover, one could argue that even the initial repetition suppression may reflect a basic form of prediction—namely, the prediction of local transitional probabilities. If so, the temporal separation of effects related to repetition and expectation may reflect different levels of predictive complexity, aligning with hierarchical predictive coding models (Garrido, Kilner, Kiebel, Stephan, et al., 2009; Kiebel et al., 2008; Todorovic & de Lange, 2012; Wacongne et al., 2012).

Across studies, predictable elements in sensory input elicited reduced neural activity. But what drives this suppression? Two major theoretical accounts have been proposed (noting that much of the supporting evidence comes from visual studies). One suggests that top-down expectations from higher-level cortical areas *filter out* predicted information, suppressing activity in early sensory regions. The alternative account proposes that prediction sharpens neural representations in early sensory areas by inhibiting neurons that do not code for expected features. This leads to a more selective population response and, consequently, a reduction in overall response (de Lange et al., 2018; Kersten et al., 2004).

Empirical evidence supports both accounts, making it difficult to determine which predominates. For example, Kok, Jehee, et al. (2012) found that expectation of a specific orientation in visual gratings suppressed activity in primary visual cortex yet improved the decoding accuracy of the grating

orientation. This finding supports the sharpening hypothesis—less activity, but more informative. Similar conclusions have been drawn in other studies (Bell et al., 2016; Yon et al., 2018). On the other hand, some research has reported the opposite pattern: suppression accompanied by *reduced* decoding performance, which would be more consistent with a filtering account (Blank & Davis, 2016; Kumar et al., 2017).

1.3.1.2 Mismatch responses

Following exposure to a regular auditory sequence, a deviation from this established pattern elicits a range of mismatch responses. These neural responses emerge from deviation onset and can extend across the following several hundred milliseconds (Southwell & Chait, 2018; Wacongne et al., 2011; Winkler, 2007). One of the most well-known responses is the mismatch negativity (MMN). As MMN is triggered by violations of expectation, it serves as an indirect marker of the brain's ability to learn and track specific regularities (Näätänen et al., 2007; Paavilainen, 2013; Winkler, 2007). MMN is typically computed by taking the difference between the neural response to standard tones and that to deviant tones, and is reflected as a negative deflection in fronto-central EEG channels, usually occurring 100–250 ms after the onset of the deviant (Winkler, 2007). One of the most common paradigms to elicit MMN is the oddball paradigm, in which a sequence of regular sounds is occasionally interrupted by a deviant. Variants of this paradigm range from simple tone repetition to more complex hierarchical patterns. One popular and well-established modification is the roving-standard paradigm, in which a specific tone feature (e.g., frequency) serves as the standard before transitioning to a new standard, allowing direct comparisons between physically identical tones presented as standard versus deviant. This design controls for stimulus-specific effects and ensures that observed MMN responses reflect contextual regularity, rather than low-level sound features.

The mismatch responses have provided compelling evidence that the auditory system can detect complex regularities beyond simple tone repetition.

Studies have shown that MMN can be elicited by deviations in frequency patterns (Bendixen & Schröger, 2008), violations of contingency rules between sound features (Paavilainen et al., 2007), and even disruptions in multisensory statistical patterns (Paraskevopoulos et al., 2018). Moreover, the magnitude and latency of the MMN are sensitive to the degree of deviation, with larger or more salient violations producing earlier and stronger responses (Fitzgerald & Todd, 2020; Garrido et al., 2013; Näätänen et al., 2007; Winkler, 2007).

Despite its widespread use, some have argued that they may reflect simple neural adaptation to repeated stimuli rather than prediction-based processes (Heilbron & Chait, 2018). However, this view has been challenged by more sophisticated paradigms. For instance, omission paradigms, in which the expected sound is omitted rather than replaced, have shown that neural responses are still evoked at the moment when a tone was expected to occur. This response cannot be attributed to the bottom-up signals, suggesting that the brain formed predictions, or expected to experience a certain sensory input (Bendixen et al., 2009; Todorovic & de Lange, 2012; Wacongne et al., 2011).

Further support comes from local-global paradigms. In a study by Wacongne et al. (2011), participants passively listened to five-tone sequences (e.g., AAAAB– AAAAB–AAAAB ...). Here, the final B represents a local deviant (violating the immediate tone pattern). Conversely, when the final B is replaced by A in an AAAAB context, A maintains local regularity but violates a global, longer-term pattern. Results showed that local deviants elicited MMN responses, whereas global deviants triggered later responses involving broader brain networks. These findings demonstrate that mismatch responses cannot be fully explained by low-level neural adaptation alone—if it were, global deviants should not elicit responses—thereby supporting the idea that mismatch responses reflect prediction-based processing.

Notably, the local-global paradigm provides compelling support for the notion of hierarchical predictive coding. In Wacongne et al. (2011), global deviants were found to activate a fronto-parietal network, whereas local

deviants did not. Combining with the finding of post-MMN response in global violation, this suggests that violations are computed at different hierarchical levels, consistent with the idea of a hierarchical predictive coding. Supporting this interpretation, an electrocorticography (ECoG) study by Dürschmid et al. (2016) similarly reported that global deviance recruited higher-order cortical regions, reinforcing the view that predictive coding operates across multiple levels of the cortical hierarchy.

Furthermore, computational modelling of MMN generation has shown that predictive coding models within hierarchical networks provide the best fit to empirical data (Garrido et al., 2008; Garrido, Kilner, Kiebel, & Friston, 2009). Within this framework, MMN is considered to be tightly associated with a concept of prediction error. Supporting this account, individuals with autism (see Section 1.2.2.1) often exhibit a reduced MMN response (Dunn et al., 2008; Gomot et al., 2011). One interpretation is that heightened sensory precision in autism may reduce the suppression of prediction errors during standard, predictable tones, resulting in a smaller difference between standard and deviant responses (Lawson et al., 2014).

Crucially, MMN can be elicited even in the absence of attention or explicit awareness of the regularity, making it a powerful tool for probing implicit prediction mechanisms in passive listeners (Bendixen et al., 2007; Bendixen & Schröger, 2008; Tivadar et al., 2021). This underscores MMN's utility as a neural marker of automatic regularity detection and highlights the remarkable capacity of the auditory system to learn statistical structure and generate predictions without conscious effort.

1.3.1.3 Decoding predicted content

Findings from MMN and repetition suppression provide indirect evidence that the brain is sensitive to statistical regularities and capable of forming predictions based on repeated exposure. A more direct approach to studying this predictive capacity involves decoding expected information from neural

activity. This line of research has been primarily pursued in the visual domain (Ekman et al., 2017; Kok, Jehee, et al., 2012; Kok et al., 2014, 2017; Summerfield & de Lange, 2014). For example, Kok et al. (2017) demonstrated that sensory expectations can pre-activate stimulus templates in the brain, rendering them decodable even before the corresponding stimulus appears. In their study, participants heard auditory cues that reliably predicted the orientation of an upcoming grating stimulus. Using multivariate decoding techniques, the authors showed that these auditory cues elicited early neural activity patterns corresponding to the expected visual orientation—prior to the actual visual input.

Demarchi et al. (2019) extended this work into the auditory domain using an omission paradigm (see above). Participants passively listened to tone sequences that varied in predictability, manipulated through the transitional probabilities of tone frequencies. By applying time-generalisation decoding approach, the researchers tested whether information about the expected tone frequency could be extracted from neural activity. They found that, in highly predictable sequences, frequency-specific information was present both before tone onset and during silent omissions. This result clearly indicates that the passive listening brain learns statistical regularities and uses them to pre-activate frequency-tuned neural ensembles.

Building on this finding, subsequent studies have explored individual variability in the strength of pre-stimulus predictive activity. Schubert et al. (2023), for instance, found that individuals with stronger anticipatory neural signals also exhibited enhanced cortical tracking of the speech envelope in a separate task, suggesting a potential functional benefit of this predictive processing. Another study focusing on individuals with tinnitus revealed altered anticipatory neural signals compared to control participants (Reisinger et al., 2024), supporting the notion that aberrant predictive processing may contribute to the experience of tinnitus (Sedley et al., 2016).

Overall, these findings demonstrate that statistical regularities are learned and encoded in anticipatory brain activity. Moreover, they suggest that such predictive processing actively shapes perception and contributes to individual differences in sensory experience—even when the external input is identical.

1.3.1.4 Sustained neural response

So far, I have been reviewing the responses represent the *outcome* of learning processes and do not directly reveal how environmental regularities are internalised over time. This raises a key question: can we observe process of the gradual formation and refinement of internal models? A growing body of research suggests that the neural mechanisms underlying the tracking of auditory statistical regularities can be studied through analyses of M/EEG sustained activity. These neural responses systematically vary with the predictability of incoming sounds, offering a more direct view of how the brain dynamically monitors the structure of its sensory environment (Barascud et al., 2016; Bianco et al., 2025; Herrmann et al., 2019, 2021; Herrmann & Johnsrude, 2018; Hu et al., 2024; Magami et al., 2025; Southwell & Chait, 2018; Zhao et al., 2025).

A commonly used paradigm in this research involves presenting participants with sequences of 50 ms tone-pips arranged in either regularly repeating (REG) or random (RND) patterns, while they passively listen. Using this paradigm, Barascud et al. (2016) investigated how sequences with varying degrees of predictability are represented in sustained MEG responses. They found that the emergence of a REG pattern was associated with a gradual increase in sustained neural activity, which plateaued after experiences a few cycles of REG (***Figure 1.1A***). In contrast, during RND sequences, neural activity plateaued at a much lower amplitude. Notably, the timing of divergence between REG and RND responses closely matched the moment when active listeners detected the regularity, suggesting that the increase in amplitude

reflects the brain's process of discovering the repeating pattern, while the plateau indicates stabilisation of the internal representation.

Moreover, the amplitude of the sustained response varied according to the predictability of the REG sequences. For example, sequences formed from larger "alphabets" (i.e., greater number of different frequencies forming the REG pattern) were harder to predict and elicited slower, more moderate increases in neural activity (***Figure 1.1A***). This pattern supports the idea that sustained response amplitude indexes the inferred predictability of ongoing auditory input.

Interestingly, when the alphabet size of the REG sequence was smaller (e.g., REG5 or REG10), the timing at which the sustained neural response diverged from that of the RND sequence closely matched the amount of information an ideal listener would need to detect the REG pattern, as predicted by a variable-order Markov model (***Figure 1.1A***; see Section 1.3.2.1 for modelling details). However, as the alphabet size increased, this correspondence weakened: the brain required more tones to detect the REG pattern than the model predicted. This discrepancy highlights the brain's limitations in discovering and maintaining representations of higher-order statistical regularities.

Similar sustained response patterns have been replicated in EEG studies (Southwell and Chait, 2018; see also Chapter 3), across different types of auditory regularities (Herrmann & Johnsrude, 2018; Sohoglu & Chait, 2016), and in stochastic sequences with varying alphabet size (Zhao et al., 2025).

Sustained neural responses are not limited to representing static auditory scenes—they also reflect the brain's dynamic tracking of changes in sound statistics over time. When the auditory environment transitions from REG to RND or from RND to REG, the sustained response changes accordingly, capturing the brain's detection of pattern violations and the discovery of new regularities (***Figure 1.1B***; Barascud et al., 2016; Bianco et al., 2025; Magami et al., 2025; Zhao et al,, 2025). The asymmetry in neural responses to these

transitions—one exhibiting a more abrupt change and the other a more gradual shift—will be further explored in Section 1.4.

Building on this, Bianco et al. (2025) extended the paradigm by replacing the RND segment with a new regular pattern. Their findings showed that the sustained response can also track the brain's re-establishment of a new regularity, highlighting its sensitivity not only to the discovery and violation of predictability, but also to the ongoing process of updating internal models in response to structured changes in the environment (***Figure 1.1C***).

While the precise computational parameters reflected in sustained response amplitude remain under active investigation (see Section 1.3.2 for further discussion), these results highlight the potential of sustained responses as a window into how the brain accumulates evidence, detects changes, and forms new predictions in real time—even in the absence of attention or explicit awareness.

It is important to note that while the broad generators of sustained responses can be identified through source localisation (see next section for details), the underlying circuit-level mechanisms remain under investigation. The response resembles a direct current (DC) shift, potentially involving potassium flux. However, to date, there are no known reports of successful measurement of this response in ECoG or animal models. Investigating this phenomenon is particularly challenging due to the low-frequency nature of the signal, which is often eliminated by the high-pass filters commonly applied in ECoG studies.

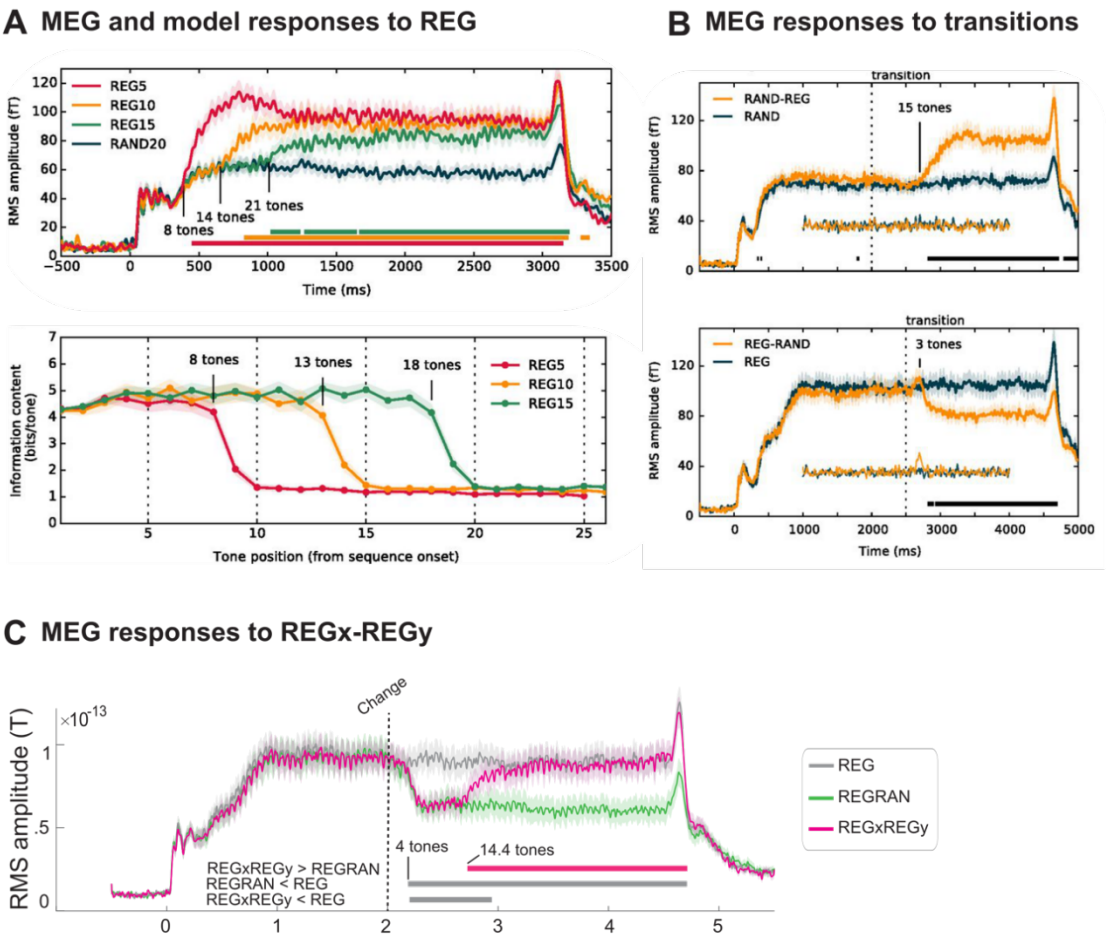


Figure 1.1 MEG responses to rapidly evolving statistical structures in sound stimuli.

[A] Top: Sustained MEG responses to regular sequences (REG) composed of different alphabet sizes (5, 10, and 15 tones), compared to random (RND) sequences as a control. Bold lines indicate time intervals where each REG condition significantly diverges from the RND condition. Numbers mark the divergence onset time for each REG condition. Bottom: Output from an ideal observer model, showing the point at which the information content (tone-evoked surprisal; see Section 1.3.2.1 for details) begins to decrease. Adapted from Barascud et al. (2016). **[B]** Sustained MEG responses during transitions from RND to REG (top) and from REG to RND (bottom), with their respective no-change controls. Bold lines indicate significant

divergence between conditions, and numbers marking the divergence onset. High-pass filtered versions of the data are plotted at the bottom. Adapted from Barascud et al. (2016). **[C]** Sustained MEG responses to REG, REG-RND, and REGx-REGy (a switch between two distinct REG sequences). Bold lines indicate the significant differences between condition pairs. Adapted from Bianco et al. (2025).

1.3.1.5 Neural networks involving regularity tracking

I have reviewed in this section that the processing of auditory regularities engages a broad network of brain regions, ranging from low-level sensory areas to higher-order cortical regions, depending on the complexity of the extracted structure and the nature of the neural activity being measured (Dürschmid et al., 2016; Kok, Jehee, et al., 2012; Wacongne et al., 2011). In line with this, Barascud et al. (2016) demonstrated that the discovery of regular patterns is supported by a distributed network involving the auditory cortex (AC), hippocampus (HC), and inferior frontal gyrus (IFG). Notably, the AC–IFG circuit has been consistently implicated in the generation of MMN, highlighting its importance in predictive auditory processing (Garrido et al., 2008; Garrido, Kilner, Kiebel, & Friston, 2007; Garrido, Kilner, Kiebel, Stephan, et al., 2007; Garrido, Kilner, Stephan, & Friston, 2009; Opitz et al., 2002; Phillips et al., 2015, 2016). Extending this, Bianco et al. (2025) showed that activity within this network dynamically fluctuates in response to changes in auditory structure: the network's engagement weakens when a regular pattern is disrupted and is re-established as a new regularity emerges. These dynamics suggest that top-down connectivity from IFG is strengthened when a predictive model is available and disrupted when the model becomes irrelevant to the current input.

The hippocampus also appears to play a key role in the rapid detection and encoding of sensory regularities (Bornstein & Daw, 2012; Schapiro et al., 2012, 2014; Turk-Browne et al., 2010). In Schapiro et al. (2012), participants were exposed to a continuous stream of fractals while performing an unrelated

task. These stimuli were organised with varying transitional probabilities (TPs), such that some pairs had high TPs while others were more weakly associated. fMRI response recorded before and after this incidental learning session revealed that exposure to these temporal regularities altered the object representation in the hippocampus: fractals with high TPs were represented more similarly than those with lower TPs, suggesting that the hippocampus encodes the statistical structure of the input. Following to that, Schapiro et al. (2014) demonstrated that individuals with hippocampal lesions showed impaired learning to temporal regularities, further confirming the hippocampus's necessity in regularity learning.

Interestingly, Bianco et al. (2025) found that when a previously established REG pattern was reinstated after a brief disruption, hippocampal responses were stronger than during the initial presentation. This finding is striking given that the sound was task-irrelevant, rapidly unfolding, and passively processed, indicating that memory traces of the auditory structure had been formed rapidly and automatically. This aligns with growing evidence that the hippocampus supports implicit memory for auditory patterns (Billig et al., 2022). Taken together, these findings suggest that the hippocampus plays a critical role not only in the rapid acquisition of sensory regularities but also in retaining and reactivating these representations to guide ongoing model construction and updates. The role of such stored representations in shaping scene prediction will be further explored in Chapter 2.

1.3.1.6 Is regularity down-weighted or up-weighted?

In Section 1.3.1.1, I introduced the phenomenon of expectation suppression, where expected stimuli tend to evoke reduced sensory responses. However, in Section 1.3.1.4, I presented findings showing that more predictable, regular sequences can actually elicit *increased* neural responses, posing an apparent contradiction. How can we reconcile these seemingly opposing effects?

One framework that offers a resolution is precision weighting (see Section 1.2.2.1). According to this account, signals that are deemed more precise (i.e., reliable) receive greater weight in neural processing. This prioritisation can amplify neural responses to deterministic sequences relative to random sequences, as observed in the REG response enhancement discussed in Section 1.3.1.4. In the predictive coding framework, precision is tightly linked with attentional gain (Feldman & Friston, 2010). Indeed, in some studies reporting enhanced neural responses to predictable stimuli, stimulus predictability covaried with attention factor (e.g., task-relevance; de Lange et al., 2018). Kok, Rahnev, et al. (2012) manipulated those factors independently to demonstrate that attention can reverse the typical suppressive effects of expectation.

However, many studies discussed in Section 1.3.1.4 (Barascud et al., 2016; Hu et al., 2024; Sohoglu & Chait, 2016; Southwell & Chait, 2018; Zhao et al., 2025) were conducted under passive listening conditions where the sounds were not behaviourally relevant. This raises the question: could regularity itself be inherently salient and thus capable of automatically attracting attention? To address this, Southwell et al. (2017) tested whether REG sequences are more salient than RND sequences, but instead found that RND sounds were actually more distracting. This result suggest that attentional mechanisms alone cannot fully explain the enhancement of responses to regular input. It remains possible, however, that other precision-related processes—distinct from both voluntary and involuntary attention—contribute to the up-weighting of regular auditory input (Southwell et al., 2017).

An alternative account proposes that the increase in sustained responses to REG may reflect heightened inhibitory activity. Since sustained responses measured with E/MEG cannot distinguish between excitation and inhibition, an increase in inhibitory processing could also produce the observed amplitude patterns (Barascud et al., 2016; Southwell et al., 2017). This inhibitory account aligns with behavioural findings that REG sequences are

easier to ignore (Southwell et al., 2017) and evoke lower arousal than RND sequences (Milne, Zhao, et al., 2021; discussed further in Chapter 4). Notably, the precision-weighting and inhibition accounts are not mutually exclusive. In fact, a growing body of evidence links precision estimation and inhibitory mechanisms (Lecaignard et al., 2022; Natan et al., 2015; Schulz et al., 2021; Yarden et al., 2022).

1.3.2 Computational models

So far, I have reviewed the neural signatures of regularity tracking in the auditory system. To deepen our understanding of the mechanisms that underlie this process, another valuable approach is the use of computational models. These models offer a framework for formalising theoretical assumptions and testing them directly against neural and behavioural data. By simulating how the brain might process and predict sensory input, computational models help uncover the underlying algorithms that support regularity tracking.

In this section, I focus on two influential models that have been extensively applied to auditory research and whose assumptions are supported by empirical findings: the **Information Dynamics of Music** (IDyOM) model and the **Dynamic Regularity Extraction** (D-REX) model.

1.3.2.1 IDyOM

The Information Dynamics of Music model (IDyOM) is a computational framework that implements a variable-order Markov model using the Prediction by Partial Matching (PPM) algorithm (Harrison et al., 2020; Pearce, 2005, 2018). Originally developed to model listeners' expectations in musical sequences, IDyOM estimates the probability of each upcoming symbol based on the sequence of preceding inputs.

IDyOM learns incrementally from symbolic sequences, such as tone-pip patterns or musical notes, and produces a conditional probability distribution for each event in the sequence, given the preceding context. These predictive

distributions are derived from n-gram models, where sequences of 'n' adjacent symbols are used to generate conditional probabilities. For example, if n = 3, the model evaluates the probability that a sequence such as “AB” will be followed by “A,” based on the relative frequency of “ABA” among all 3-grams beginning with “AB” in the learned dataset. Crucially, IDyOM integrates probability estimates from multiple n-gram models of varying orders (i.e., different values of *n;* ***Figure 1.2A***).

Using the resulting probability distributions, IDyOM calculates the information content (IC) for each symbol in a sequence. IC is defined as the negative log probability of a symbol occurring, conditioned on the portion of the sequence heard so far, and serves as a measure of the symbol’s unexpectedness. Higher IC values reflect greater deviation from expectation. IC provides a dynamic readout of the model’s ongoing adaptation to the statistical structure of the input.

IDyOM has often been used as an ideal-observer model, simulating a theoretically optimal learner with perfect memory. As such, it provides a benchmark against which human behavioural and neural responses can be evaluated. Deviations from the model's predictions offer insights into cognitive constraints and mechanisms that shape real-world auditory perception.

For instance, Di Liberto et al. (2020) used IDyOM to model melodic expectations in natural music. They demonstrated that fluctuations in model-derived expectations significantly predicted cortical responses during music listening. Their results provide compelling evidence that listeners' melodic expectations can be explained by the statistical learning model.

Barascud et al. (2016) also used IDyOM to benchmark the earliest point at which listeners could detect the emerging regularity in tone-pip sequences. The comparison between IC estimates and sustained MEG activity revealed striking alignment between them, suggesting that the inferred predictability of the auditory scene was mirrored in neural dynamics. This analysis not only

demonstrated where the brain behaves like an ideal observer, but also highlighted its limitations, such as memory constraints, which caused deviations from the model's predictions (see Section 1.3.1.4). More explanations and the actual implementation of the model are provided in Chapter 2 and 3.

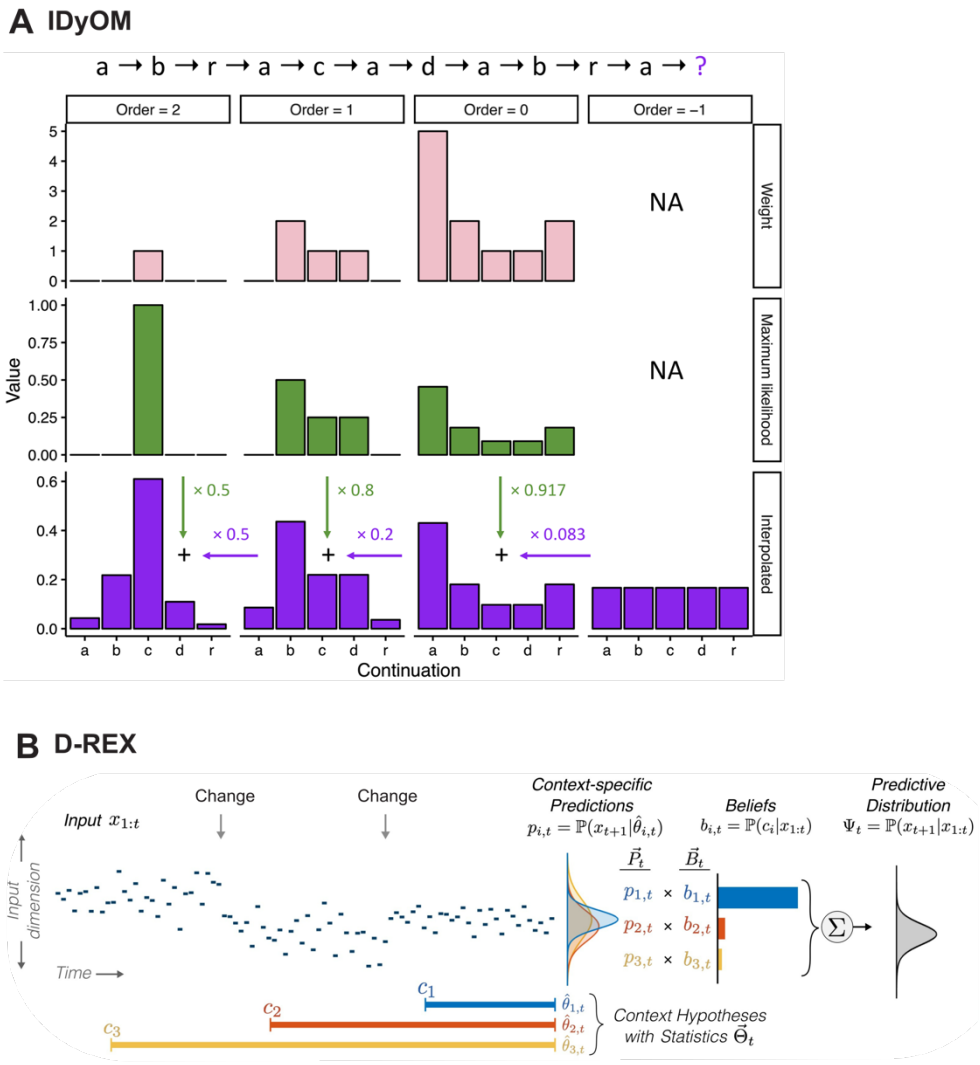


Figure 1.2 Model illustrations.

[A] Illustration of how different n-gram models with different orders are integrated in IDyOM. In this example, the order bound is 2, with five possible symbols in the sequence, and the task is to predict the next

symbol (‘?’). The top row shows the weights—i.e., the frequency of each symbol—for each order. For example, the order-0 model reflects the overall frequency of each symbol encountered so far. Order-1 model reflects the frequency of each symbol following ‘a’. The middle row displays the corresponding maximum-likelihood distributions, obtained by normalising the weights. The bottom row shows the interpolated distribution, which combines the maximum-likelihood distribution of the current order with the interpolated distribution of the next lower order. For details on how the weight of each distribution is computed, see Harrison et al. (2020). Adapted from Harrison et al. (2020). **[B]** Illustration of how D-REX combines predictions from multiple context hypotheses to account for an unknown change point. For each context, predictions are generated based on statistics accumulated within that context. These predictions are weighted by their corresponding beliefs, and the weighted combination of all context-specific hypotheses forms the overall predictive distribution for the next input. Adapted from Skerritt-Davis and Elhilali (2021).

1.3.2.2 D-REX

Dynamic Regularity Extraction (D-REX) model is a computational framework grounded in predictive coding theory based in Bayesian inference (Skerritt-Davis & Elhilali, 2018, 2021a, 2021b). D-REX operates on an ‘observe–predict–update’ loop: after each new observation in a sequence, the model generates a predictive distribution for the next observation, based on the previously encountered inputs. One of the core features of D-REX is its ability to monitor for potential changes in the underlying generative structure of the sequence. This capability is crucial in volatile environments, where maintaining an accurate internal model requires not only prediction, but also sensitivity to sudden shifts—*change points*—in ongoing sequences (see Section 1.4 for further discussion).

To accommodate this, D-REX simultaneously maintains multiple hypotheses about the current state of the environment (context hypotheses; see ***Figure 1.2B***), each assumes a different potential change point. For example, one hypothesis may assume that no change has occurred, incorporating the entire sequence history to form predictions. Another might assume that a change occurred just one tone ago and ignore earlier inputs to focus only on recent, relevant evidence for the current environment. These hypotheses are each weighted by a belief, or predictive probability of each hypothesis. The final prediction is then generated by combining these weighted hypotheses into a single predictive distribution.

In addition to its predictive output, D-REX also provides a measure of surprisal—the degree of mismatch between the predicted and actual input—and a signal of *precision (*inferred reliability*)*, defined as the inverse variance of the predictive distribution. It also estimates the likelihood that a change point has occurred at any given moment.

D-REX has been successfully applied to a range of auditory perception tasks (Skerritt-Davis & Elhilali, 2018, 2021b; Zhao et al., 2025). For example, in tasks requiring detection of changes in rapid auditory sequences, the model has demonstrated strong alignment with listener performance (Skerritt-Davis & Elhilali, 2018). Furthermore, Zhao et al. (2025) recently extended the paradigm introduced by Barascud et al. (2016) by employing stochastic tone-pip sequences to examine whether sustained neural responses reflect change point estimation or the precision of predictions. Their findings revealed that the dynamics of the sustained responses were best accounted for by the *precision* computed by the D-REX model. This suggests that sustained neural response reflects the precision, or confidence, assigned to predicted sensory input while passively tracking sound sequences. More explanations and the actual implementation of the model are provided in Chapter 3.

Taken together, these findings provide converging evidence for the auditory system's remarkable sensitivity to regularities and its capacity to generate and continuously update internal models of the sensory environment. This continuous model updating—based on accumulated input history—is highly effective in stable, well-controlled contexts such as laboratory settings. However, real-world environments are often volatile, rendering simple cumulative updating suboptimal. In such settings, the ability to detect changes in the underlying statistical structure becomes critical. The next section explores the importance of this change detection process and how the brain may implement it.

1.4 Importance of change point estimation

Imagine walking through a familiar city: your expectations, shaped by past experiences, guide your navigation. But upon entering a forest, those city-based expectations no longer apply. Detecting such an abrupt change should prompt the brain to discard outdated priors and construct a new model tailored to the new context. This prevents interference from irrelevant memories and supports efficient adaptation to the new environment. However, if it is unclear whether you've entered a forest or simply a park within the city, abruptly resetting your internal model may be risky. In such cases, a more adaptive strategy is to retain the existing model while gradually updating priors by accumulating additional evidence. This section examines how, and under what conditions, the brain detects change points in the environment and adapts its information processing strategies accordingly.

1.4.1 Change point detection in decision making

How the brain predicts upcoming events while accounting for potential changes in the environment has been a major topic in the field of decision-making. Imagine you have a favourite café. One day, however, the coffee tastes

terrible. Now you're faced with a decision: was this an indication of a fundamental change in the café's quality—suggesting you should stop going—or was it just a random anomaly you can ignore? This scenario captures a core challenge in decision-making: accurately inferring whether an unexpected event reflects a genuine change in the environment (a change point), or just random noise. A wide range of studies has shown that people are surprisingly adept at detecting such change points in dynamic environments (Boubenec et al., 2017; Glaze et al., 2015; McGuire et al., 2014; Nassar et al., 2010, 2012; Skerritt-Davis & Elhilali, 2018).

For example, Nassar et al. (2010) used a task in which participants had to predict a number drawn from a Gaussian distribution. The mean of this distribution represented the current state of the environment, and its standard deviation represented environmental noise. Occasionally, the mean would abruptly shift, introducing a change point. Participants made predictions of the number, saw the true number, and adjust their prediction to minimise prediction errors. A key concept here is the learning rate—a parameter that determines the extent to which new information influences the internal model. A learning rate of zero implies complete reliance on prior beliefs, whereas a learning rate of one implies full updating based solely on the new observation. The study found that participants' learning rates increased immediately following a change point, indicating that they were sensitive to change points and adapted their internal models accordingly by reducing reliance on outdated information.

1.4.2 How can the brain achieve change point detection?

Detecting change points in dynamic environments poses a significant challenge because, as illustrated in the café example, observers are never explicitly told when a change has occurred. One theoretically optimal solution is to track all possible change points and generate predictions for each hypothesis. This is the strategy employed by D-REX, and many other 'full-Bayesian' models which maintains multiple concurrent beliefs about when a

change might have happened and integrates them to form a final prediction (Adams & MacKay, 2007; Skerritt-Davis & Elhilali, 2018).

However, this approach is computationally demanding. As the sequence progresses, the number of potential change points increases, requiring a growing memory load and extensive computations to evaluate every hypothesis. In practice, such a model would demand infinite memory and processing capacity—an unrealistic requirement for biological systems.

To better understand the brain's strategy, researchers have tested simplified models that relax the computational demands of ideal observer frameworks (McGuire et al., 2014; Nassar et al., 2010, 2012; Wilson et al., 2013). For instance, Nassar et al. (2010) compared two models: a full-Bayesian model and a reduced-Bayesian model to explain participants' behaviour introduced above. The full-Bayesian approach keeps a probability distribution over all possible change point locations. In contrast, the reduced-Bayesian model considers only two possibilities at each trial: whether the new observation comes from the same distribution (no change) or from a new one (a change has occurred). Surprisingly, the reduced-Bayesian model performed comparably to the full-Bayesian model in capturing human behaviour, suggesting that full-Bayesian computations may not always be necessary to explain observed behavioural patterns.

It is important to note that the studies discussed above primarily stem from the sequential decision-making literature, which typically involves active engagement and focused attention. In contrast, much less is known about how change points are detected automatically during the processing of rapidly evolving auditory scenes—despite the fact that such detections are crucial for accurately tracking dynamic sequences in sound. How does the brain manage to detect changes under these conditions, often without active attention allocation?

One key factor that supports change detection is environmental volatility—the expected frequency with which the underlying statistical properties or rules of an environment changes. In highly volatile contexts, prediction errors are more likely to signal genuine change points. In contrast, under stable conditions, similar deviations are more often attributed to noise. Empirical studies have demonstrated that people are sensitive to environmental volatility, and when volatility is high, individuals tend to increase their learning rate, or sensitivity to new information relative to prior beliefs (Behrens et al., 2007; Glaze et al., 2015; Piray & Daw, 2024; see also Chapter 2 for further discussion).

Another key factor is the magnitude of the prediction error scaled by the variability (noise) of the generative distribution, or precision. When an environment is noisy (i.e., the underlying distribution is wide), prediction errors are more likely to be tolerated as plausible outcomes of the same distribution. However, in less noisy (narrower) environments, the same error magnitude is more likely to indicate a shift to a new distribution. In other words, whether an error is interpreted as a change point depends on both its magnitude and the expected variability of the distribution. This relationship has been demonstrated in several studies (McGuire et al., 2014; Nassar et al., 2010; Piray & Daw, 2024). The next section introduces this second point more closely as a potential neural proxy for change point detection in the passive listening brain.

1.4.3 Unexpected uncertainty as a proxy for change point

As reviewed above, ideal observer models offer a powerful solution to change point detection by maintaining and updating multiple hypotheses about when changes might occur. However, this approach demands extensive computational resources and is unlikely to be the brain’s default strategy. Instead, a more plausible mechanism may involve tracking a signal known as unexpected uncertainty.

The brain constantly encounters uncertainty in sensory input, which can be broadly divided into *expected* and *unexpected* uncertainties (Yu & Dayan, 2005). Expected uncertainty reflects the variability the brain anticipates based on environmental stochasticity, such as background noise in the busy café. Unexpected uncertainty, on the other hand, arises when input deviates beyond what is expected, potentially indicating a fundamental change in the environment, or change point (Bland & Schaefer, 2012; Yu & Dayan, 2005). Indeed, unexpected uncertainty and change points evokes similar neural responses.

Once a change point occurs, the brain must rapidly discard outdated beliefs and prioritise incoming sensory data to adapt to the environment (Skerritt-Davis & Elhilali, 2018). One system thought to mediate this shift is the locus coeruleus–norepinephrine (LC-NE) system. Norepinephrine (NE), released from the locus coeruleus (LC), plays a crucial role in network reset and in rebalancing top-down and bottom-up information processing (Aston-Jones & Cohen, 2005; Bouret & Sara, 2005; Joshi et al., 2016). This system enables the brain to redirect attention to salient cues and enhances sensitivity to new evidence (Devauges & Sara, 1990; Jepma & Nieuwenhuis, 2011; Lawson et al., 2017, 2021; Nassar et al., 2012; Sara, 2009; Sara & Bouret, 2012).

For example, Devauges and Sara (1990) pharmacologically enhanced LC-NE activity in rats performing a maze task. When the task rules were unexpectedly altered, rats with increased LC-NE activation adapted more rapidly to the new condition. In a separate experiment, the same manipulation led rats to spend more time exploring novel stimuli. These findings suggest that the LC-NE system supports model reset and the initiation of new evidence gathering.

Further evidence comes from a study by Lawson et al. (2021), in which human participants categorised images as faces or houses. Each image was preceded by a sound cue indicating the likely category of the upcoming visual stimulus, allowing participants to form expectations that sped up their

responses. When participants received a NE blocker, their response times improved even further—likely due to a reduced influence of bottom-up sensory processing, increasing reliance on top-down expectations. However, when the cue–category relationship was covertly reversed—effectively introducing a change point—the NE-blockage delayed adaptation to the new rule, suggesting impaired belief updating by utilising bottom-up signals. This finding reinforces the idea that NE plays a central role in facilitating belief update in response to unexpected environmental change.

The LC-NE system is also activated by experiences of unexpected uncertainty and may induce internal model reset and initiation of bottom-up processing (Basgol et al., 2025; O'Reilly, 2013; Payzan-LeNestour et al., 2013; Sara & Bouret, 2012; Yu & Dayan, 2005; Zhao, Chait, et al., 2019). Supporting evidence comes from studies using pupil dilation response (PDR) as a proxy for LC-NE activity.

For instance, in Nassar et al. (2012), participants predicted numbers sampled from a Gaussian distribution whose mean occasionally shifted. Results showed that PDR increased following surprising outcomes, and this increase positively correlated with both the change point probability and the learning rate. Interestingly, the study extended this finding to task-irrelevant events that nonetheless evoked unexpected uncertainty. Occasionally, a surprising auditory change—irrelevant to the task and not informative of the generative model—was introduced. These unexpected sounds elicited PDR and boosted learning rates in the subsequent primary task. This suggests that the brain may treat unexpected uncertainty, even if it is task-irrelevant, as signals of change point, triggering a model reset even when no actual change point is present (Nassar et al., 2012; Yu, 2012). The influence of task-irrelevant unexpected uncertainty on performance in an attended task will be explored in greater detail in Chapter 4.

A complementary line of evidence comes from Zhao, Chait, et al. (2019), who measured PDR while participants passively listened to auditory sequences that transitioned either from regular to random (REG-RAN) or from random to regular (RAN-REG). Although both transitions involved a change in statistical structure, the experienced abruptness of the change differed due to the predictability of the pre-transition sequence. In the REG-RAN condition, the highly predictable regular sequence generated precise expectations, making the transition to randomness feel like a sharp and unanticipated deviation—i.e., unexpected uncertainty. In contrast, the RAN-REG transition emerged from an unpredictable context, where higher expected uncertainty made the emergence of structure feel more gradual and expected. Critically, PDR was observed only in the REG-RAN transition, suggesting that the brain selectively tracks unexpected uncertainty as a cue for potential change points (Basgol et al., (2025) replicated this finding). However, when participants attended actively to the RAN-REG transitions, PDR was evoked also to this transition direction. This finding implies that LC-NE-mediated model resets are preferentially engaged by unexpected uncertainty when attention is limited, but that gradual changes can also elicit LC-NE responses when the system is actively engaged and computational resources are available.

A similar distinction between abrupt and gradual changes, as observed in Zhao, Chait, et al. (2019), is also evident in sustained neural responses (Barascud et al., 2016; Bianco et al., 2025; Zhao et al., 2025). For example, Barascud et al. (2016) showed that REG-RAN transitions elicited a sharp reduction in sustained neural activity, whereas RAN-REG transitions produced a more gradual (***Figure 1.1B***). These two distinct responses may also reflect different computational strategies, a full reset of the internal model in response to unexpected changes, and continuous updating in the face of slowly emerging structure. Chapter 2 and 3 investigate this hypothesis further.

In summary, these findings suggest that the brain may rely on unexpected uncertainty as a heuristic for change detection, especially when

scene tracking is not the primary task. This strategy likely confers computational efficiency, enabling the brain to remain sensitive to salient shifts without the burden of continuously maintaining and updating a full set of hypotheses. However, definitive evidence that unexpected uncertainty triggers model reset and enhances bottom-up processing during passive listening remains limited. This question will be further explored in Chapters 2 and 3.

1.5 Aim of this project

In everyday life, we are constantly surrounded by dynamically fluctuating background sounds that often fall outside the focus of our attention. Yet, these task-irrelevant sounds are far from inert. As reviewed above, the auditory system is highly adept at tracking regularities in the environment, even in the absence of immediate behavioural relevance. While our attention tends to prioritise goal-directed tasks, the brain concurrently processes background regularities, which may shape internal models of the world, modulate arousal, and influence how we explore and interpret incoming information. However, our current understanding of these processes primarily stems from studies using simple stimuli that do not reflect the complexity of real-world auditory scenes.

This PhD thesis aimed to deepen our understanding of regularity processing in dynamically changing environments and to explore how such computations influence broader brain functions. How do we navigate an uncertain auditory world, and in turn, how does this shape the way we perceive, respond to, and act within it?

Chapter 2 investigates how the brain uses prior experience to inform ongoing sound sequence processing. While prediction relies on past sensory input, the relevance of that past information can vary greatly in everyday environments. This experiment examined whether the passively listening brain can flexibly evaluate the utility of past information and determine when it should

be integrated into ongoing predictive models. To address this question, the study combined EEG recording with computational modelling.

Chapter 3 explores how the predictability of a preceding sequence affects the efficiency of regularity detection in a subsequent auditory scene. As previously discussed, regularity detection is a fundamental process in auditory perception, but it is often studied in isolation. In real-world contexts, however, regularities arise within continuous streams of sensory input, and the brain typically approaches new input with an already-formed predictive model rather than from a neutral starting point. This chapter sought to embed regularity detection within a more ecologically valid framework by examining how prior context influences the processing of emerging regularities. EEG and computational modelling were employed to investigate these dynamics.

Chapter 4 turns to the question of how the automatic processing of background sound sequences impacts performance on an attended task. Typically, we process environmental sound while our primary attention is allocated elsewhere. While previous research has demonstrated that background sound regularities are continuously tracked, little is known about how such processes influence the execution of an unrelated, attention-focused task. This study examined the influence of dynamic auditory backgrounds on memory encoding, drawing on event boundary theory—a framework describing how we segment and store continuous experience. Using an audiovisual behavioural task combined with skin conductance measurements, this experiment investigated whether task-irrelevant background sound could modulate memory for concurrent visual events.

These studies build on previous findings related to sustained neural responses (Barascud et al., 2016; Bianco et al., 2025; Hu et al., 2024; Zhao, Chait, et al., 2019; Zhao et al., 2025). To place the research questions outlined above in a more concrete context, I briefly summarise the relevant literature introduced in this chapter. ***Figure 1.3*** illustrates the sustained neural response

to a sound sequence that transition from one regularity to another (REGx–REGy) reported in Bianco et al. (2025). The sustained neural response has been hypothesised to reflect the precision of the ongoing sequence: it increases as the brain detects REGx, drops upon violation of REGx, and then gradually recovers as REGy is discovered—eventually stabilising at a sustained amplitude following this second discovery (Barascud et al., 2016; Bianco et al., 2025; Magami et al., 2025; Zhao et al., 2025). The sharp drop in sustained response following REG violation is accompanied by activation of the pupil-linked LC-NE system (Basgol et al., 2025; Zhao, Chait, et al., 2019), and has been interpreted as reflecting a reset of the brain's predictive model. In contrast, transitions from a random sequence to a regular one do not elicit abrupt shifts in sustained activity or engage the LC-NE system, suggesting that regularity discovery unfolds more gradually through evidence accumulation.

Chapter 2 focuses on this post-discovery neural response, examining whether and how prior auditory information influences on the ongoing scene predictions reflected in sustained neural activity. Chapter 3 investigates the discovery trajectory of REGy, testing whether its emergence is shaped by the statistical structure of preceding sounds. Finally, Chapter 4 contrasts two types of pattern transitions: REG violation (an abrupt change) and REG discovery (a gradual change). Here, I assess how these distinct transition dynamics influence the encoding of concurrently presented visual information.

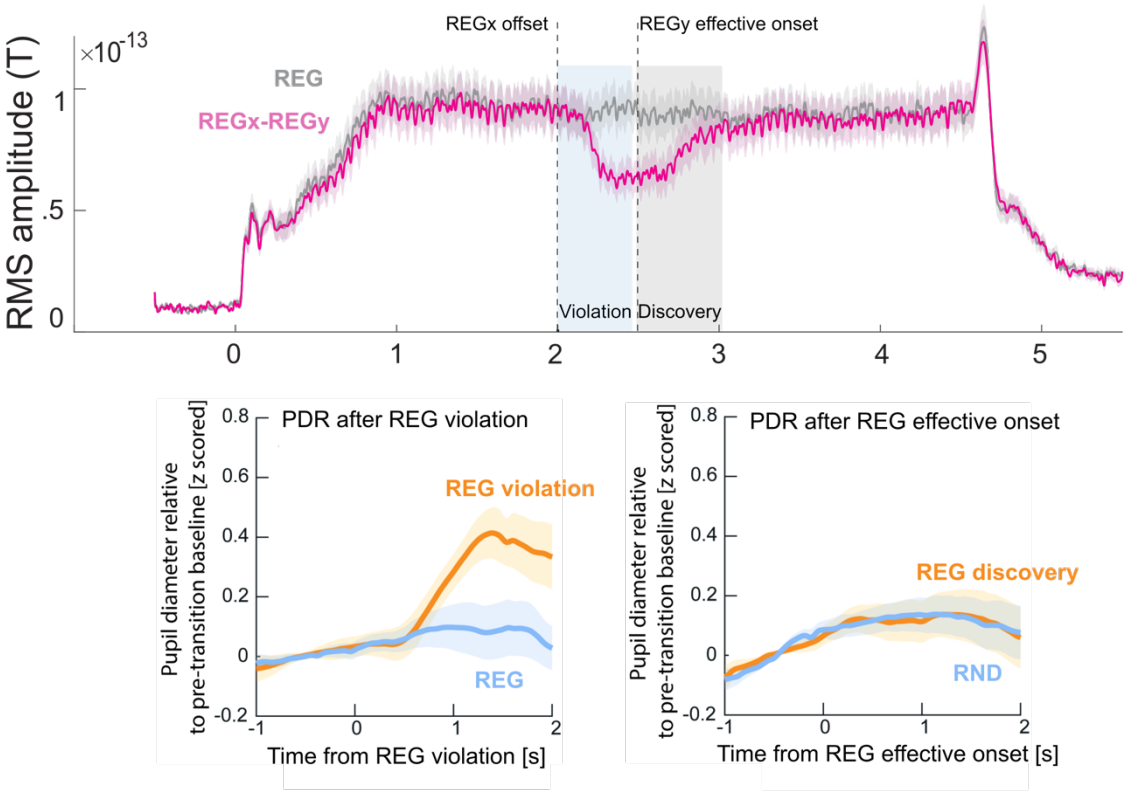


Figure 1.3 Summary of current understanding of sustained neural responses.

Top: Sustained MEG response to regular sequences (REG) and to sequences with a transition between two distinct regularities (REGx–REGy). The first dotted line marks the offset of REGx, while the second marks the effective transition point to REGy, defined as the onset of its second cycle. The blue shaded area highlights neural responses associated with the violation of REGx, and the grey shaded area indicates responses related to the discovery of REGy. Adapted from Bianco et al. (2025). Bottom: Pupil dilation responses to REG violation (transition from REG to RND, left) and REG discovery (transition from RND to REG, right). Adapted from Zhao, Chait et al. (2019).

Throughout this thesis, I employ rapidly evolving tone-pip sequences arranged in either regular (REG) or random (RND) patterns. Each tone pip has a duration of 50 ms, a timescale relevant for auditory perception and corresponding to the shortest known information integration window in the auditory cortex (Norman-Haignere et al., 2022; Saberi & Perrott, 1999). The rapid pace of these sequences makes it unlikely that listeners can consciously track the patterns, enabling investigation of pre-attentive neural mechanisms. Further, all sound sequences introduced in this thesis are drawn from the same frequency pool and carefully controlled for specific frequency composition. This design ensures that any observed effects can be attributed to the statistical structure of the sequence, rather than to low-level acoustic differences. An additional advantage of using such stimuli is the absence of semantic or emotional content, which are common confounds in naturalistic sounds such as music or speech-like artificial grammars. This reduces individual variability linked to experience and affect. Most importantly, these stimuli are highly quantifiable and well-suited for computational modelling.

2. Chapter 2: The Effect of Previously Encountered Sensory Information on Neural Representations of Predictability: Evidence from Human EEG

2.1 Summary

Accumulating evidence suggests that the brain continuously monitors the predictability of rapidly evolving sound sequences, even when they are not behaviourally relevant. An increasing body of empirical evidence links sustained tonic M/EEG activity to evidence accumulation and tracking the predictability, or inferred precision, of the auditory stimulus. However, it remains unclear whether, and how, this process depends on auditory contextual memory. The present EEG study examined neural responses to sound sequences across two experiments, and compared them to predictions from ideal observer models with varying memory spans. Stimuli were sequences of 50 ms long tone-pips. In Experiment 1 (N=26; both sexes), a regularly repeating sequence of 10 tones (REG) transitioned directly to a different regular sequence (REGxREGy). In Experiment 2 (N=28; both sexes), the same regular sequence was repeated after an intervening random segment (REGxINTREGx). Results from Experiment 2 revealed that the inferred predictability of the resumed REGx pattern was influenced by the preceding INT tones, even several seconds after they ended, indicating that the brain retains contextual memory over time. In contrast, neural responses in Experiment 1 were best explained by models with minimal memory. This dissociation implies that the brain can dynamically adjust its strategy based on inferred environmental structure—resetting context when interruptions signal change, and preserving context when patterns are likely to resume.

This chapter has been adapted from a submitted paper: Magami K, Bianco R, Hall E, Pearce M, Chait M. 2025. The effect of previously

encountered sensory information on neural representations of predictability: evidence from human EEG. bioRxiv. https://doi.org/10.1101/2025.05.27.656332.

2.2 Introduction

As reviewed in Chapter 1, human brain is remarkably sensitive to the statistical regularities ubiquitously present in our surroundings (Arnal & Giraud, 2012; Bendixen, 2014; Bendixen et al., 2012; de Lange et al., 2018; Maheu et al., 2019; Press et al., 2020; Willmore & King, 2023; Winkler et al., 2009). A large body of research has demonstrated that observers can automatically acquire complex statistics from sensory inputs, including auditory, visual, and multimodal streams (Boubenec et al., 2017; Conway & Christiansen, 2005; Demarchi et al., 2019; Fiser & Aslin, 2001; Garrido et al., 2013; Horváth et al., 2001; Saffran et al., 1999; Stefanics et al., 2014; Turk-Browne et al., 2009; Wacongne et al., 2011). This computational ability is critical for generating predictions about the environment (Bendixen, 2014; Bendixen et al., 2012; de Lange et al., 2018; Friston, 2005; Press et al., 2020; Winkler et al., 2009), which allows the brain to optimise behaviour by efficient allocation of cognitive and neural resources, supporting adaptive responses to incoming events (Bendixen et al., 2012; Boubenec et al., 2017; Bouwkamp et al., 2025; Kok, Jehee, et al., 2012; Nobre et al., 2007; Southwell & Chait, 2018; Yon et al., 2018).

An increasingly well-supported observation is that the neural mechanisms underlying the tracking of auditory statistical regularities can be studied through analyses of M/EEG sustained activity. These neural responses systematically vary with the predictability of sequential inputs, providing a direct window into how the brain monitors and adapts to environmental statistics (Barascud et al., 2016; Herrmann et al., 2019, 2021; Herrmann & Johnsrude, 2018; Hu et al., 2024; Southwell & Chait, 2018; Zhao et al., 2025).

Experiments using rapidly evolving auditory sequences have progressively revealed how the auditory system processes and accumulates

statistical information about the acoustic environment. In the standard paradigm (e.g., Barascud et al., 2016), participants passively listen to tone sequences that transition between regular (REG) frequency patterns and random (RND) patterns. These sequences elicit a sustained neural response that dynamically tracks the structure of the auditory input (Barascud et al., 2016; Hu et al., 2024; Southwell et al., 2017; Zhao et al., 2025). Specifically, the emergence of a REG pattern is associated with a gradual increase in sustained neural activity, which plateaus as the regularity becomes established, suggesting that the brain has stabilised a representation of the repeating structure **(Figure *1.1*A).** Notably, longer and more complex patterns result in slower and more moderate amplitude increases, indicating limitations in the brain's ability to discover and maintain representations of higher-order statistical regularities. Upon transition from REG to RND, the sustained response drops sharply and then settles into a lower, stable level—interpreted as reflecting the low predictability of random sequences. Zhao et al. (2025) extended these findings to stochastic sequences consisting of RND patterns with different predictability. These rises and drops in the sustained response align with predictions from computational ideal observer models (Harrison et al., 2020; Pearce, 2005; Skerritt-Davis & Elhilali, 2018, 2021a), which quantify information content (IC; how surprising a given tone is based on prior exposure) or precision (inferred reliability of the predictive distribution; Yon & Frith, 2021), providing support for the hypothesis that the sustained response represents a mechanism that tracks predictability within the unfolding signal. However, it remains unclear how the brain determines the context or reference frame, whether derived from immediate sensory input or retrieved from longer-term memory, against which this predictability is assessed.

Commonly used modelling measures such as information content (e.g. as used in Harrison et al., 2020; Pearce, 2005), or precision (e.g. as used in Zhao et al. 2025), quantify the expectedness of an event given a particular context of previously encountered events stored in memory. In theory—drawing from Bayesian change-point estimation models (Adams & MacKay, 2007;

Fearnhead & Liu, 2007; Wilson et al., 2010)—ideal observers should dynamically evaluate the relevance of a given context and determine how much of it to incorporate when constructing predictive distributions (Glaze et al., 2015; Nassar et al., 2010; Skerritt-Davis & Elhilali, 2018, 2021a; Wilson et al., 2013). However, whether and how the sustained response, as a proxy for predictability processing, depends on experienced auditory events remains unexplored. Addressing these questions is crucial for understanding how past experiences are leveraged to represent the predictability of a given event.

Bianco et al. (2025) showed that REG patterns were recognised more quickly by the brain (as indicated by the MEG sustained response) when they were re-introduced following a scene interruption than when initially presented. This indicates the presence of an automatic memory store that carries a representation of the pattern across the interruption. More broadly, this finding suggests that by manipulating the information encountered by listeners and measuring its effects on the sustained response, it is possible to gain insight into what information is being stored and the conditions under which it is utilised. Here, I ask: Will the sustained response to a regular pattern be influenced by a listener's prior experience with past information? This is tested by comparing two situations that differ in the relevance of prior experience: one in which a regularity is learned and then replaced by a new one—the prior experience is no longer relevant (**Experiment 1**), and another in which a regularity is learned, interrupted by a random tone sequence whose length is varied systematically, and then resumed—the prior experience is relevant and could be carried over (**Experiment 2**).

2.3 Experiment 1

This experiment examines changes in the EEG sustained response triggered by transitions between two distinct regular (REG) patterns—REGx to REGy – compared with a continuation of REGx (***Figure 2.1A***). A similar comparison was made in one of the experimental conditions reported by Bianco

et al. (2025) using MEG. Here, I replicate that approach using EEG to justify the use of EEG in the extension reported in Experiment 2.

To inform the interpretation of the data, I use IDyOM, which implements a variable-order Markov model based on the Prediction by Partial Matching algorithm (Harrison et al., 2020; Pearce, 2005). The model has been extensively and successfully used to account for regularity processing in artificial sequences, such as those used in the current study (Barascud et al., 2016; Bianco et al., 2020, 2025; Harrison et al., 2020), as well as in more naturalistic musical settings (Cheung et al., 2019, 2023; Di Liberto et al., 2020; Kern et al., 2022; Quiroga-Martinez et al., 2021).

Starting with a null model, IDyOM learns incrementally based on the unfolding tone sequence and uses its learned model to generate a conditional probability distribution for each tone given the preceding tones. ***Figure 2.1B*** shows model predictions. To simulate the availability of different amounts of contextual information for REGy pattern detection, I varied the duration of the input sequence—referred to as the "pre-training window"—the model was trained on before the transition to REGy. In the simulations shown in ***Figure 2.1B,*** this window ranged from just a few seconds to 240 trials. Model predictions were always based on the full context available up to that point (i.e., all prior input; see figure legend). This approach enabled systematic manipulation of the model's memory content to examine how varying levels of prior information influence its output. The model quantifies the information content (IC) of each tone —reflecting the surprise elicited by that tone given the preceding context. I compare a *context incorporating model* that retains increasingly long spans of past input (from a few seconds to the entire experiment) with a *reset model* that clears its memory upon detecting a deviant tone—the first tone in REGy that violates expectations based on REGx.

All models show a gradual decrease in IC during REGx as the pattern is learned. This decrease occurs at different rates depending on the pre-training

window; models with longer pre-training windows exhibit greater variability across trials (indicated by larger error bars) due to cumulative influences from prior tones. At the transition to REGy, all models show a sharp increase in IC, corresponding to the surprise elicited by an unpredictable tone. IC then remains high for a period before gradually reducing again, indicating that the new regularity (REGy) is being learned. The duration of this learning period varies across models: models with a shorter pre-training window (e.g., Model 1.2) take longer to adapt (reflected by a slower decrease in IC), as existing memory content of REGx interferes with the encoding of the new pattern. Conversely, models with a longer pre-training window (e.g., Model 1.4) or where training is reset (e.g., Model 2) exhibit a more rapid decrease in IC, as the representation of REGy is less affected by prior memory of REGx.

This example also illustrates how the difference in IC between REGy and the non-changing REG control condition is modulated by the pre-training window. When the pre-training window is short, the difference in IC is consistently large, as the memory of REGx strongly influences the encoding of REGy. However, as the model's pre-training window increases in length, this difference diminishes due to memory saturation: previously encountered patterns interfere with both REGx and REGy representations. In the *reset* model, the IC difference is also low, reflecting a complete lack of memory competition.

As discussed, the M/EEG sustained response is thought to reflect neural tracking of sequence predictability. If the brain represents sequence information similarly to the model, neural activity would be expected to mirror the dynamics of IC.

2.3.1 Methods

2.3.1.1 Stimuli

The stimuli (***Figure 2.1A****)* were 3500 ms long sequences composed of 50 ms tone pips (5 ms raised cosine ramps; 70 tone pips in total). REG sequences were generated by randomly selecting 10 frequencies from a pool of

20 logarithmically spaced values between 222 and 2000 Hz without replacement, and this sequence was cycled to create a regularly repeating pattern. For the REGxREGy sequence, two distinct REG sequences were generated: the first REG pattern (REGx) lasted 2 s, and the second (REGy) lasted 1.5 s. REGx was formed by randomly selecting 10 frequencies from the pool without replacement, and the remaining 10 frequencies were used to form REGy (***Figure 2.1A***). A unique sound sequence was generated for each trial and participant. The inter-stimulus interval (ISI) was jittered between 2.5 and 3 s.

2.3.1.2 Procedure

These data were collected as part of a separate study (reported in Chapter 3), that contained other stimuli (presented in a separate block).

Participants were seated in an acoustically shielded room (IAC triple-walled sound attenuating booth). They listened to auditory stimuli while engaging in a decoy visual task, presented on a computer screen located about 90 cm away. The visual task consisted of sequentially presented triplets of photographs of landscapes, and participants were instructed to press a key when the first and third images were the same (occurring in 40% of trials). Feedback regarding the number of hits, misses, and false alarms for the visual task was provided at the end of each block. The duration of the image presentation was jittered between 2 and 5 s, and images were cross faded to avoid abrupt visual transients. The timing of image presentation was not correlated with that of the auditory stimuli.

Overall, 120 sound stimuli were presented for each of the two sound conditions (REG, REGxREGy). These stimuli were presented randomly and arranged in 4 blocks. Sounds were presented diotically through headphones (3A Insert Earphone, 3M) via a Fireface UC sound card (RME) at a comfortable listening level (adjusted by each participant). Stimulus presentation was controlled with the Psychtoolbox package (Psychophysics Toolbox Version 3) in MATLAB (2019b The MathWorks, Inc.).

2.3.1.3 Recording and data processing

EEG signals were recorded using a Biosemi system (Biosemi Active Two AD-box ADC-17, Biosemi, Netherlands) from 64 electrodes at a sampling rate of 2048 Hz. Recording was restarted at the beginning of each block. For data analysis, the Fieldtrip (http://www.fieldtriptoolbox.org/) toolbox for MATLAB (2018a, MathWorks) was used.

The recorded data were down-sampled to 256 Hz, low-pass filtered at 30 Hz (two-pass, Butterworth, 5th-order) and detrended by a 1st-order polynomial. The data were divided into epochs of 6 s, from 1 s pre-stimulus onset to 1.5 s post-stimulus offset. The epochs were then baseline-corrected relative to the pre-onset interval (-0.5 s to 0 s relative to the sound onset). Outlier epochs and channels were removed by visual inspection, resulting in the removal of an average of 4.24 % of epochs and 0.9 channels per participant. De-noising source separation (DSS; De Cheveigné & Parra, 2014; De Cheveigné & Simon, 2008) analysis was then applied to each subject's data across all conditions to maximise reproducibility across trials (over the interval of 0 s to 4 s relative to sound onset). For each participant, the first three DSS components were retained and projected back into sensor space. Finally, the data were re-referenced to the average of all channels, and the averages over epochs for each channel, condition and subject were calculated.

To quantify the effects, the most auditory-responsive channels were selected: for each participant, the N1 component (negative event-related potential happening at around 100 ms post-stimulus onset) of the sound onset response was identified from the averaged data. At the peak of the N1, the 5 channels showing the most positive activity and the 5 channels showing the most negative activity were considered to best reflect the brain's auditory-related activity. In the figures below, I quantify the instantaneous power of the brain response by computing the RMS (root mean square) across these channels, following a similar approach in other works (Barascud et al., 2016; Southwell et al., 2017; Zhao et al., 2025). The RMS reflects the instantaneous

power of the brain response regardless of polarity. Field maps at relevant time points are also provided.

2.3.1.4 Statistical analysis

To statistically evaluate the effect of interruption, the differences between sound conditions were calculated for each participant. This difference was then subjected to bootstrap resampling (Efron & Tibshirani, 1994). The difference between conditions was considered significant if the proportion of bootstrap iterations falling above or below zero exceeded 99% ($p<.01$) for more than 8 adjacent samples (Barascud et al., 2016).

2.3.1.5 Participants

Twenty-eight paid participants participated in Experiment 1. All reported no history of hearing or neurological disorders. Two participants were excluded due to exceptionally noisy EEG data. Data from the remaining twenty-six participants (19 females; average age 24.81, ± 4.20) were used for analyses. All experimental procedures were approved by the research ethics committee of University College London, and written informed consent was obtained from each participant.

2.3.2 Results and discussion

2.3.2.1 The EEG sustained response tracks regularity discovery and violation

The group averaged responses for the two conditions (REG, REGxREGy) are shown in ***Figure 2.1C***. Overall, findings from Bianco et al. (2025) were successfully replicated with EEG.

The brain response exhibited an N1 peak at around 100 ms post-onset, then increased its amplitude until it reached a plateau before the end of the 2nd cycle of the REG sequence. This sustained response pattern aligns with previous literature and is thought to reflect a rapid, automatic process of regularity detection (Barascud et al., 2016; Herrmann et al., 2019, 2021; Herrmann & Johnsrude, 2018; Hu et al., 2024; Southwell et al., 2017). Following the emergence of the REGy pattern, the sustained response rapidly dropped in

amplitude, persisted at a low level (whilst the new REG pattern was being discovered) and then returned to the pre-transition level. To analyse the difference in the post-transition responses between conditions, data were baseline-corrected relative to the pre-transition window (1.5-2 s post-onset; ***Figure 2.1C***, right). Bootstrap resampling revealed a significant difference between the amplitudes of REG (control) and REGxREGy, starting from 220 ms (~5 tones) after the transition, consistent with the timing shown in the MEG data from Bianco et al. (2025).

As noted previously (Barascud et al., 2016; Bianco et al., 2025), unlike during regularity discovery, the EEG response latency here diverges from model predictions, which show a spike in IC immediately following the first tone that violates the REG pattern. Several factors could account for this divergence. One possibility is that the delay reflects a circuit-related delay in encoding the violation of the REG pattern. Alternatively, it might reflect a "wait-and-see" period, during which the system accumulates information about the scene change before responding. Indeed, Bianco et al. (2025) demonstrated that this latency is not fixed but scales with sequence information content (tone-pip duration), challenging the idea of a simple refractory period.

Following the abrupt drop in the sustained response, levels remained low for a period before rising again. The difference between conditions disappeared at 800 ms post-interruption (16 tones), at which point the response to REGy returned to the levels of the no change control condition (REG). Overall, these patterns indicate that the EEG sustained response dynamically tracks the brain’s process of discovering predictability, detecting its violation, and then fully re-establishing a new regularity.

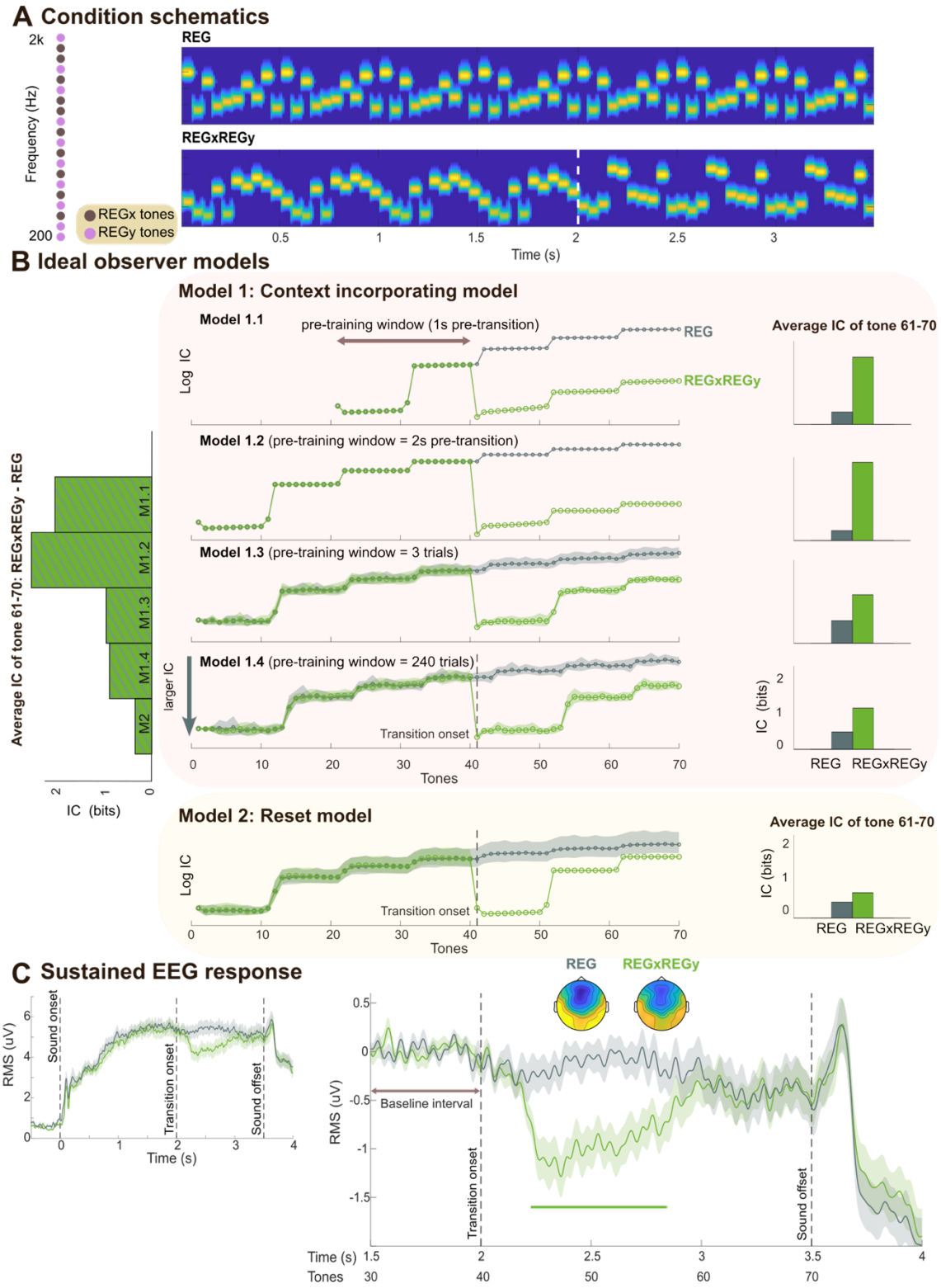

A Condition schematics
2k
Frequency (Hz)
200
REGx tones
REGy tones
REG
REGxREGy
0.5
1
1.5
2
2.5
3
Time (s)
B Ideal observer models
Model 1: Context incorporating model
Model 1.1
pre-training window (1s pre-transition)
REG
REGxREGy
Log IC
Model 1.2 (pre-training window = 2s pre-transition)
Model 1.3 (pre-training window = 3 trials)
Model 1.4 (pre-training window = 240 trials)
larger IC
Transition onset
0
10
20
30
40
50
60
70
Tones
Average IC of tone 61-70
IC (bits)
REG
REGxREGy
Average IC of tone 61-70: REGxREGy - REG
M1.1
M1.2
M1.3
M1.4
M2
Model 2: Reset model
C Sustained EEG response
RMS (uV)
Sound onset
Transition onset
Sound offset
Time (s)
Baseline interval
Time (s)
Tones

Figure 2.1 Experiment 1: stimuli, model simulations, and EEG results.

[A] Left: Schematic illustration of the frequency selection method in Experiment 1, with each frequency represented as a circle. In this example, the brown frequencies were allocated to REGx and the pink to REGy. Right: Spectrograms depicting example stimuli for each condition. The dashed line marks the onset of REGy. **[B]** Model simulations. The model was implemented by using the "new_ppm_simple" function from the ppm R package, available on GitHub (https://github.com/pmcharrison/ppm). All parameters were kept with their default settings, as described in the repository documentation. Middle: Information content (IC; log transformed) computed from variants of the IDyOM model, each incorporating different memory constraints. The y-axis is inverted (bottom = higher IC). The REG condition is shown in grey; REGxREGy condition is shown in green. For each condition, data are averaged over trials, with shaded areas representing twice the standard deviation (STDEV). Models vary by the duration of the "pre-training window". Model 1.1 is pre-trained on 2 cycles of REGx (indicated by the brown arrow). Model 1.2 is pre-trained on 4 cycles of REGx. Model 1.3 is pre-trained over 3 trials. Model 1.4 is pre-trained over all 240 trials. Model 2 is reset upon pattern interruption, resulting in a pre-training window of length zero. All models estimate variable-order conditional probabilities for the next tone given the immediately preceding sequence of tones. The stimulus context over which the model learns representations of statistical structure that inform its conditional probabilistic predictions consists of the pre-training window and all tones experienced up to the time of prediction. The context varies between models, for example when predicting tone 50, the context is: Model 1.1, tones 20-49 of the current sequence; Model 1.2, tones 1-49 of the current sequence; Model 1.3,

the three preceding trials plus tones 1-49 of the current sequence; Model 1.4, the 240 preceding trials plus tones 1-49 of the current sequence; Model 2, tones 41-49 of the current sequence. Right: Raw (non-log-transformed) IC values averaged over the last REGy cycle (tone 61 to 70; corresponding to 3-3.5 s). Left: IC differences (between REGxREGy and REG computed over tone 61-70) across all five models. **[C]** EEG data. Left: Group-averaged brain responses (RMS over 10 most responsive auditory channels; see Methods). Shading indicates twice the SEM (computed via bootstrap resampling, 1000 iterations). Data are baseline-corrected relative to the 0.5-second pre-onset window. Right: The same data but baseline-corrected using the 1.5–2 s pre-transition window. Significant differences ($p<.01$) between conditions are indicated by the horizontal bold line. Scalp topographies are based on activity averaged over the time window of significant differences (2.2-2.8 s relative to stimulus onset); the colour ranges from -4 to 4 uV.

2.3.2.2 Reconciling differences between modelling and the EEG response

There are notable differences between the EEG responses and the model's behaviour. For instance, as previously noted, the model exhibits an immediate response to the transition from REGx to REGy, whereas brain responses show a delay of about five tones.

A key point of divergence lies in how the model handles multiple cycles of regularity. Even in the control condition (no change, REG), the model continues to refine its representation of REG with each successive cycle. In contrast, the EEG sustained response to REG plateaus after approximately two cycles, indicating that the brain’s representation stabilises relatively quickly.

As a result, in the model, REGy never reaches the same representational strength as REG in the control condition, since REG continues to be refined indefinitely. However, EEG data show that the transition from REGx to REGy

leads to a return to the same level of sustained activity observed for REG within about one second of REGy onset. This discrepancy suggests that aspects of evidence accumulation—or more generally, auditory processing—that shape brain responses are not fully captured by the model.

Despite these differences, the models most consistent with the EEG findings are those in which the post-transition difference between REG and REGy is minimal—that is, models in which REGx and REGy do not strongly compete in memory. Such models typically either have a long pre-training window (e.g., Model 1.4) or are reset at the point of transition (Model 2), enabling a rapid reinstatement of a REGx-like response to REGy.

To further refine this interpretation, Experiment 2 asked how prior context affects the 'rediscovery' of a previously experienced regularity. To address this, responses to an identical REG pattern were examined while systematically varying the immediately preceding context.

2.4 Experiment 2

This experiment investigated the EEG sustained response evoked by an ongoing regular (REG) pattern occasionally interrupted partway. A stimulus set (**Figure *2.2A***) was employed in which 25% of the trials consisted of a regularly repeating sequence of tones. In the remaining trials, the regular pattern was interrupted by the insertion of 1, 3, or 5 novel tones (referred to as conditions INT1, INT3, and INT5, respectively) after which the original REG pattern resumed. I asked how this interruption would affect the representation of REG, with a specific focus on the speed at which the regularity was re-discovered and the post-interruption sustained response.

As in Experiment 1, the hypothesis was constrained using IDyOM (**Figure *2.2B***). In this case, the stimuli consisted of a continuous REG pattern interspersed with occasional deviant tones. As a result, the IC differences between conditions are markedly smaller than those observed in Experiment 1

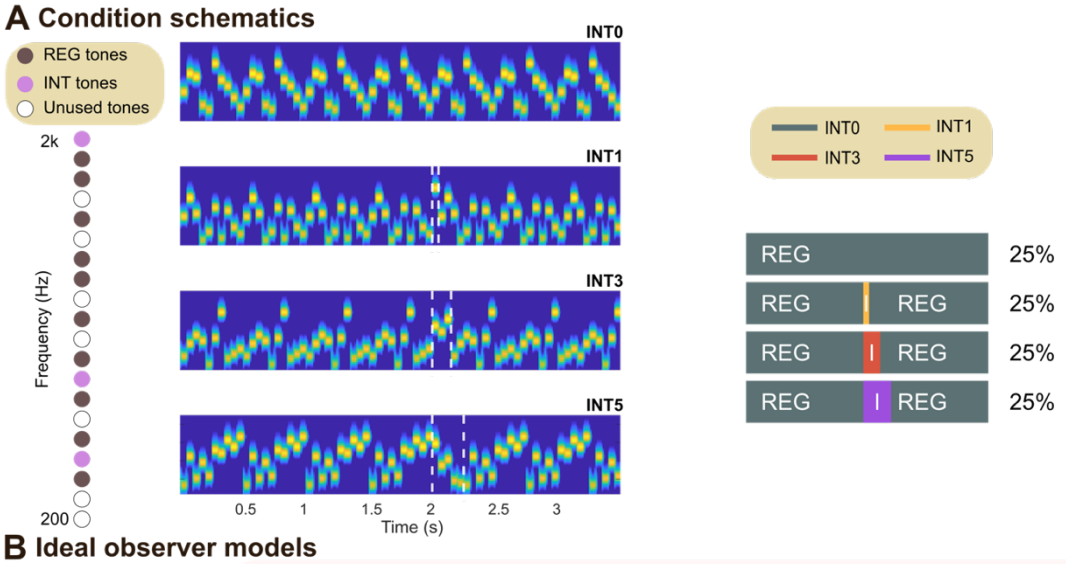
(**Figure *2.1B***). Nevertheless, the overall dynamics are consistent with those reported previously.

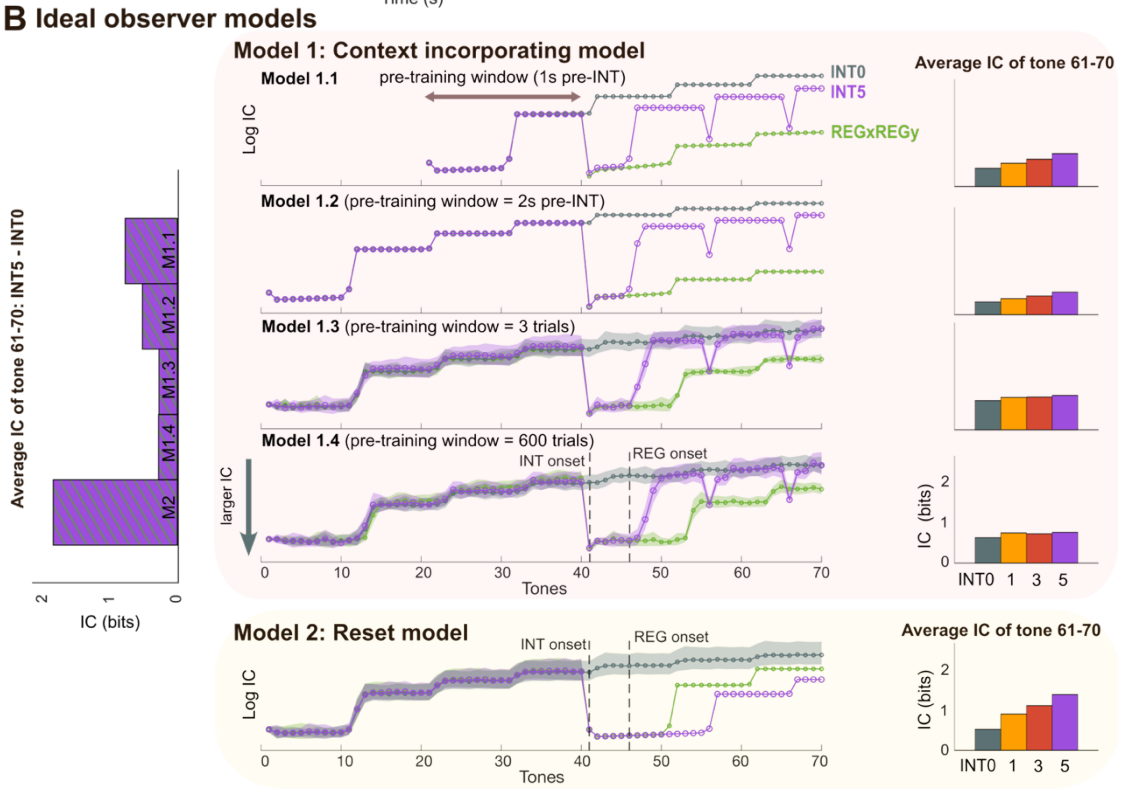
Importantly, the post-interruption behaviour of the *context incorporating* model reveals two key phenomena: (1) Despite the post-interruption sequences being structurally identical across the INT conditions, IC levels remain distinct between them (**Figure *2.2C***). This occurs because the model incorporates the interruption tones into its predictive framework, increasing baseline uncertainty. (2) The model exhibits “phantom” IC spikes, reflecting an expectation for the interruption to recur. This behaviour arises because the model lacks the capacity to infer higher-order rules, such as the one-time occurrence of interruptions and the guaranteed resumption of the pre-interruption pattern. Overall, the model's behaviour is dictated by its perfect memory of all prior experiences, with every past observation—regardless of its present relevance—being weighted equally. This includes the singleton interruptions, which continue to influence the model's present IC estimates. This pattern is largely preserved across models with different pre-training window lengths, though models with longer pre-training show less pronounced differences in post-interruption IC (due to memory saturation; as discussed in Experiment 1, above).

For a model whose memory is reset at the interruption (Model 2), IC differences between conditions are also present because the ‘post-interruption world’ contains different numbers of unique elements for each interruption condition. The interruption tones are incorporated into model predictions, thereby decreasing baseline predictability. This model does not display phantom spikes, because its memory does not contain the previously experienced REG and its transition to the interruption tones.

Another difference between the models concerns the speed at which the REG pattern is re-discovered, reflected in the timing of the decrease in IC following the interruption. The *context-incorporating* models exhibit a rapid re-discovery of REG, whereas in the *reset* model, this process is slower due to the unavailability of pre-interruption memory.

Building on these insights, this experiment examines whether passively listening participants exposed to these sequences will mirror model behaviour. Specifically, the investigation focused on whether transient disruptions affect subsequent representations of regularity in a manner comparable to the model.


A Condition schematics
REG tones
INT tones
Unused tones
2k
200
Frequency (Hz)
INT0
INT1
INT3
INT5
0.5
1
1.5
2
2.5
3
Time (s)
INT0
INT1
INT3
INT5
REG
25%
REG
REG
25%
REG
REG
REG
25%
REG
REG
REG
25%


B Ideal observer models
Model 1: Context incorporating model
Model 1.1
pre-training window (1s pre-INT)
INT0
INT5
REGxREGy
Log IC
Model 1.2 (pre-training window = 2s pre-INT)
Model 1.3 (pre-training window = 3 trials)
Model 1.4 (pre-training window = 600 trials)
INT onset
REG onset
larger IC
0
10
20
30
40
50
60
70
Tones
Average IC of tone 61-70
IC (bits)
INT0
1
3
5
Average IC of tone 61-70: INT5 - INT0
M1.1
M1.2
M1.3
M1.4
M2
2
1
0
IC (bits)
Model 2: Reset model
INT onset
REG onset
Log IC
Tones
Average IC of tone 61-70
IC (bits)
INT0
1
3
5

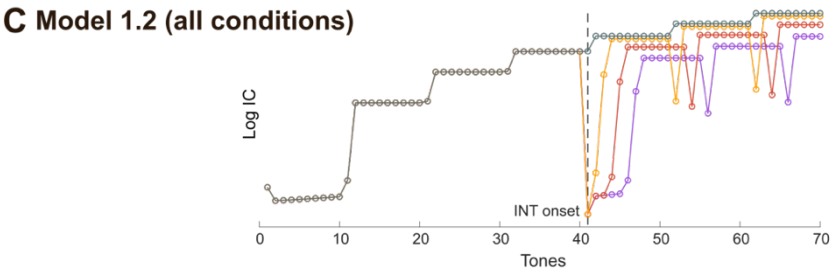

C Model 1.2 (all conditions)
Log IC
INT onset
0
10
20
30
40
50
60
70
Tones

Figure 2.2 Experiment 2: stimuli and model simulations.

[A] Left: Schematic of the frequency selection method in Experiment 2, with each frequency represented as a circle. The brown circles represent frequencies randomly chosen for REG; the pink circles represent tones chosen for the interruption tones (INT3 here). White circles denote unused frequencies in this trial. Middle: Spectrograms showing example stimuli for each condition. The white dashed box highlights the INT tones. The REG sequences before and after the INT tone follow an identical pattern. Right: Design schematics illustrating the stimulus sequences for each condition. 'I' indicates the interruption tones. **[B]** Model simulations. Middle: IC values (log-transformed) computed from variations of the IDyOM model, each with different memory constraints (as detailed in ***Figure 2.1***), plotted for the INT0 and INT5 conditions from Experiment 2, and REGxREGy condition from Experiment 1. For each condition, data are averaged over trials, with shaded areas representing twice the standard deviation (STDEV). The y-axis is inverted (bottom = higher IC). Right: Raw (non-log-transformed) IC values averaged over the last REG cycle (tone 61 to 70; corresponding to 3-3.5 s). Left: IC differences (between INT5 and INT0 computed over tone 61-70) across all five models. **[C]** IC values for all four conditions, computed using Model 1.2.

2.4.1 Methods

2.4.1.1 Stimuli

The stimuli were 3500 ms long sequences of 50 ms tone pips (5 ms raised cosine ramps; 70 tone pips in total). Tone frequencies were drawn from a pool of 20 logarithmically spaced values between 222 and 2000 Hz. Each stimulus comprised a sequence of regularly repeating tones (REG), generated in the same manner as in Experiment 1 (**Figure *2.2A***). In 25% of trials, the REG pattern continued with no interruption (INT0). In the remaining trials, an

interruption in the form of 1, 3, or 5 new tones was introduced at 2000 ms post-onset (following 4 cycles of REG). These conditions will be referred to as INT1, INT3 and INT5, respectively. The frequencies of INT tones were randomly selected without replacement from the pool of remaining frequencies not used to form the REG sequence. Following INT, the original REG pattern was re-started. The duration of this remaining portion varied across conditions (1500 ms, 1450 ms, 1350 ms, and 1250 ms for INT0, 1, 3, and 5 conditions, respectively), ensuring that the overall tone number remained fixed at 70 tones. The ISI was jittered between 2.5-3 s. A unique sound sequence was generated for each trial and participant.

2.4.1.2 Procedure

General procedures were identical to those in Experiment 1. Overall, 600 sound stimuli were presented (150 stimuli per condition; in random order). The session was divided into 5 blocks, each approximately 10 min long. Participants were allowed a short rest between blocks.

2.4.1.3 Recording and data processing

General protocols were identical to those described in Experiment 1. On average, 1.47% of epochs were removed as outliers, along with 0.5 channels per participant. For the detailed comparisons of RMS values between conditions, two different baseline correction time windows were applied to the output RMS. For the comparison of the post-interruption neural response, baseline correction was applied at the time window before the interruption onset (1.5 s to 2 s post-onset). To compare the timing where the neural response after interruption tones stabilises, baseline correction was applied at a different time window (3 s to 3.3 s post-onset). Additionally, post-interruption neural responses were compared across INT conditions (INT1, INT3, and INT5) by subtracting the control condition (INT0) from each INT condition, followed by baseline correction in the 1.5-2 s post-onset window.

To uncover activity potentially masked by the slow DC changes, the same analysis was performed on high-pass filtered data at 2 Hz (two-pass,

Butterworth, 4th-order) with baseline correction applied just before the onset of the interruption (1.8 s to 2 s post-onset). DSS was applied to the data around the interruption tone (1.5 s to 4 s post-onset), and 2 components of the DSS outputs were retained for the data representation. Given the particular interest in the mismatch negativity (MMN)-like response to the pattern interruption, electrodes that best reflected the MMN response were selected. To do this, the data were averaged across all conditions across all participants and the 10 electrodes with the most negative activation at the typical MMN response time (150 ms to 200 ms post-interruption-onset) were selected.

To examine the possible presence of the “phantom” interruption peaks, the high-pass filtered data at 2 Hz were analysed by applying DSS separately to each experimental condition (2 s to 4 s post-onset) and 2 components were extracted for the data representation. Analysing each condition separately was necessary because the “phantom” peaks occur at a different latency in each condition (tone 52 and 62 in INT1, 54 and 64 in INT3, and 56 and 66 in INT5; ***Figure 2.2B, C***). The output data were then segmented into 600 ms epochs (from 400 ms before the model-based peak timing to 200 ms after the peak timing) to limit the analysis at around the model-inferred peak locations. The initial 200 ms of each epoch was used for baseline correction, and the responses from 10 channels (same as those used in the RMS calculation) were averaged.

2.4.1.4 Statistical analysis

To statistically evaluate the effect of interruption, the differences between sound conditions (INT0, INT1, INT3, and INT5) were calculated for each participant. This difference was then subjected to bootstrap resampling (Efron & Tibshirani, 1994). The difference between conditions was considered significant if the proportion of bootstrap iterations falling above or below zero exceeded 99% ($p<.01$) for more than 8 adjacent samples (Barascud et al., 2016).

For the bootstrap analysis on data baseline-corrected to 3-3.3 s post-onset, the final significance point was defined as the moment when the neural

response stabilised. To assess whether this timing differed across conditions, the bootstrap analysis was repeated for each condition pair (INT0 vs. INT1, INT0 vs. INT3, and INT0 vs. INT5). Specifically, 1000 iterations of bootstrap resampling were performed per pair and the last significant data point within the interval from INT offset to 3 s (the onset of the baseline correction window) was identified in each iteration.

2.4.1.5 Participants

Thirty paid participants participated in Experiment 2. All reported no history of hearing or neurological disorders. Two participants were excluded due to exceptionally noisy EEG data. Data from the remaining twenty-eight participants (22 females; average age 23.4, ± 3.41) were used for analyses. All experimental procedures were approved by the research ethics committee of University College London, and written informed consent was obtained from each participant.

2.4.2 Results and discussion

2.4.2.1 All interruption conditions elicit early MMN-like responses

Sensitivity to the interruption was evaluated by analysing the response at the transition. To isolate the MMN-like response which was expected to be evoked by the INT (deviant) tones (**Figure *2.3***), the EEG data were high-pass filtered at 2 Hz and averaged across trials for each condition (as detailed in the Methods section). Indeed, the response is not visible in the non-high-pass-filtered data; see **Figure *2.4***. Bootstrap resampling revealed significant deflection in the INT1, INT3, and INT5 conditions relative to the INT0 condition (**Figure *2.3***), with latencies emerging between 70-100 ms post interruption onset. Notably, this latency and the corresponding topography (**Figure *2.3***; bottom) are consistent with those commonly associated with the MMN response (Winkler, 2007). Overall, this suggests that the interruption was similarly detected by the brain in all conditions.

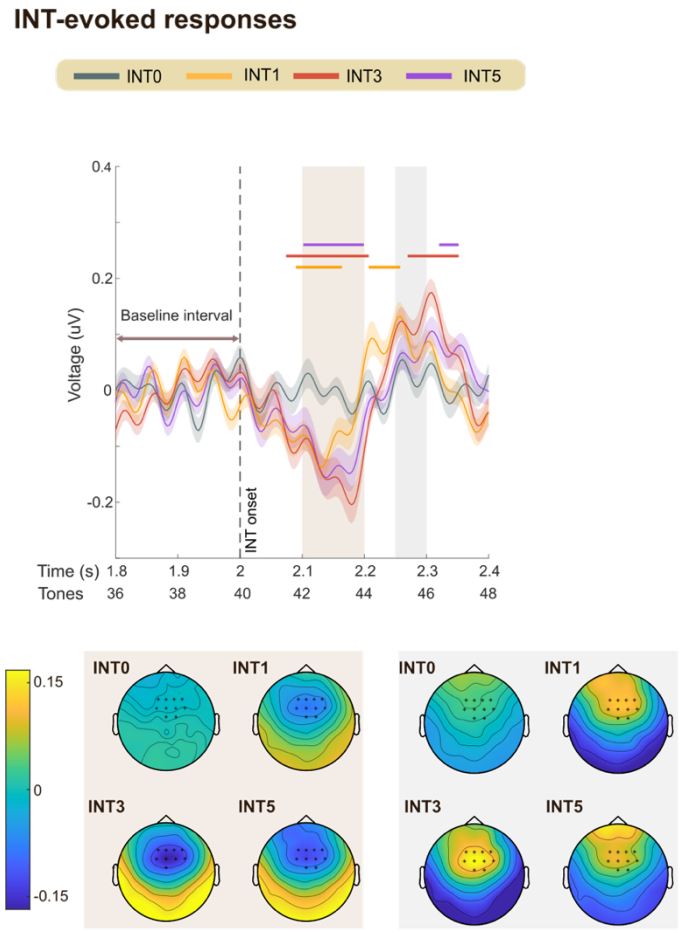


Figure 2.3 Experiment 2: INT evoked deviance response.

High-pass filtered mean EEG data, averaged over 10 channels (indicated in the scalp topographies). Shaded areas represent twice the SEM. Significant differences ($p<.01$) between INT0 and interruption conditions (INT1, INT3, INT5) are indicated by the horizontal lines above the EEG traces. Scalp topographies, calculated for two time windows (2.1-2.2 s and 2.25-2.3 s), are shown at the bottom.

2.4.2.2 The EEG sustained response tracks the dynamics of sequence IC

For each participant and condition, the RMS over 10 selected channels (detailed in the Methods section) was calculated at each time point on the non-

high-pass filtered data. The group averaged RMS amplitude for the four conditions (INT0, INT1, INT3, and INT5) is shown in **Figure *2.4A***. The general trajectory mirrored the pattern observed in Experiment 1: the sustained response increased and plateaued as the brain adapted to the REG pattern, dropped in amplitude following the INT tones, and then recovered (but not fully to baseline) as the brain re-engaged with the REG pattern. This trajectory aligns with the information content patterns predicted by the IDyOM models.

Following the REG interruption, the sustained response dropped rapidly. To analyse differences in post-interruption responses across conditions, the data were baseline-corrected relative to the pre-interruption window (1.5-2 s post-onset; **Figure *2.4B***). Bootstrap resampling (see Methods) revealed a significant difference between the control condition (INT0) and the interruption conditions (INT1, INT3, and INT5), starting at 187 ms (~4 tones) after the interruption onset. This finding is consistent with previous observations (Barascud et al., 2016; Bianco et al., 2025), including Experiment 1 in this study.

As noted previously, one possibility is that the delay reflects a fixed refractory period after the MMN-like response or some other circuit-related delay in encoding the violation of the REG pattern. Alternatively, it might reflect a "wait-and-see" period. If the 4-tone latency reflects a period of assessment—during which the system evaluates whether the REG violation is a spurious event or indicative of a consistent stimulus change—no interruption response would be expected in INT1, but a larger response would be expected in INT5. This was partially observed: while all interruption conditions exhibited a similar latency for the sustained response drop, the trough was deeper for INT3 and INT5 than for INT1 (**Figure *2.4D***).

2.4.2.3 The EEG sustained response indicates a memory trace for REG post interruption

To assess the time required to re-learn the REG pattern—reflected in the recovery of the sustained response— the data were baseline-corrected relative to the post-recovery window (3-3.3 s post-onset; indicated in **Figure *2.4C***).

Bootstrap resampling (see Methods) identified time points where responses to the interruption conditions (INT1, INT3, and INT5) remained significantly below the control (INT0) condition. Amplitude recovery was defined as the latest time point where this significant difference was observed. This occurred approximately 216 ms (~4 tones), 202 ms (~4 tones), and 285 ms (~5 tones) after the offset of the final interruption tone in the INT1, INT3, and INT5 conditions, respectively.

Under perfect memory conditions—as seen in the model dynamics—the timing of REG re-discovery following the interruption should be the same across all INT conditions, once the duration of the interruption is accounted for (i.e., subtracting 1, 3, or 5 tones, respectively; e.g., see **Figure *2.2C***). In contrast, the results show a longer latency following INT5 compared to INT1 and INT3. Bootstrap resampling confirmed a consistent difference between the INT1/INT3 and INT5 conditions (**Figure *2.4C***), suggesting that INT5 requires one additional tone to re-establish the REG pattern after REG is reintroduced. This may reflect neural memory constraints that limit the speed of pattern re-learning.

Critically, and notwithstanding the differences between conditions highlighted above, the observed recovery times were consistently shorter than a regularity cycle (i.e., <10 tones) and faster than model predictions for the discovery of a new REG pattern (i.e., 1 cycle + ~5 tones, as observed in Experiment 1). This indicates that, despite the interruption, the brain retained a memory of the REG pattern, enabling faster re-discovery (see also Bianco et al., 2025).

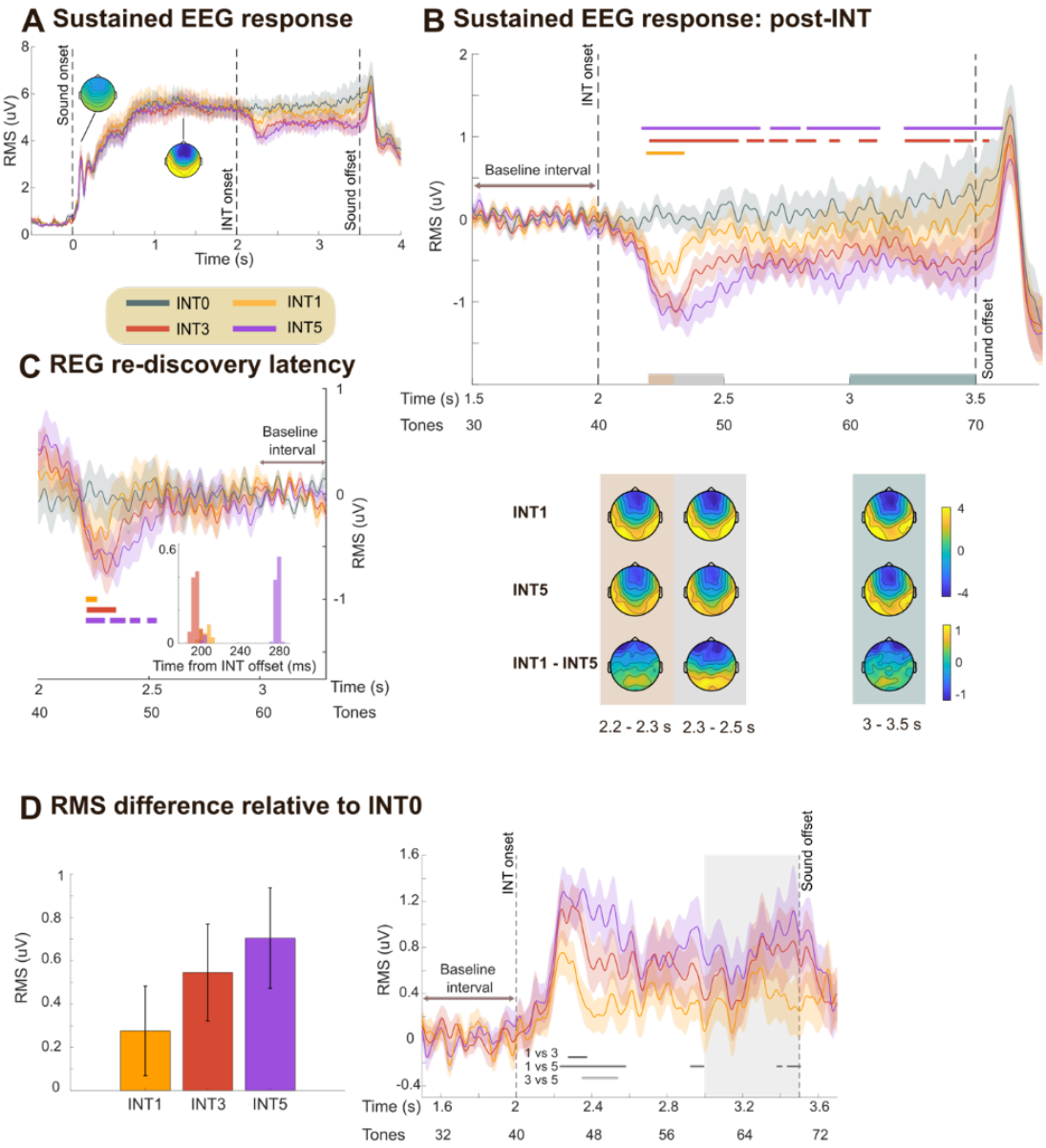


Figure 2.4 Experiment 2: sustained response dynamics.

[A] Group-averaged RMS of brain responses. Shaded areas represent twice the SEM. Data are baseline-corrected to the -0.5-0 s pre-onset window. Scalp topographies illustrate two response phases: N1 component (80-150 ms post-sound onset) and the sustained response (1-2 s post-sound onset); the colour ranges from -4 to 4 uV. **[B]** Same data as in [A] but baseline-corrected to the pre-interruption window (1.5-2 s). Significant differences (p<.01) between INT0 and INT1, INT3,

INT5 are indicated by bold horizontal lines above the EEG traces. Scalp topographies are provided for three time windows: 2.2-2.3 s, 2.3-2.5 s, and 3-3.5 s relative to sound onset. **[C]** Same data as in [A], baseline-corrected to 3-3.3 s. Significant differences ($p<.01$) between INT0 and INT1, INT3, INT5 are indicated by bold lines below the EEG traces. The histogram (inset) shows the latencies associated with REG re-discovery. The results demonstrate delayed rediscovery of REG in the INT5 condition. **[D]** Right: EEG data (RMS) for each of the interruption conditions after subtracting the INT0 condition, baseline-corrected within the pre-transition window (1.5-2 s). Grey lines beneath the traces mark significant differences ($p<.01$) between INT conditions (INT1 vs. INT3, INT1 vs. INT5, INT3 vs. INT5). Left: Same data averaged over the 3-3.5 s time window (grey shading). Error bars represent SEM.

2.4.2.4 Persistent post-interruption sustained response differences between conditions

In contrast to the results in Experiment 1, after the interruption, persistent differences in the sustained response between INT 1, 3, 5 and INT0 were observed; **Figure *2.4B*** indicates that sustained responses did not return to the pre-interruption baseline in INT conditions. Given that the amplitude of the sustained response is hypothesised to reflect the brain's representation of the predictability of unfolding sounds, this reduced amplitude suggests a decrease in inferred predictability with exposure to a greater number of INT tones, similar to that observed in modelling.

One salient feature in the model is the expectation of "phantom" interruption events at the onset of every regularity cycle following the pattern interruption (tone 52 and 62 in INT1, 54 and 64 in INT3, and 56 and 66 in INT5, as shown in **Figure *2.2B and C***). To examine potential EEG correlates of these events, the data were high-pass filtered and RMS over 10 selected channels

was calculated (see Methods). However, this analysis did not yield consisted EEG parallels. It remains possible that the noisy nature of EEG signals obscured them and that the fluctuations in the time domain (e.g. see **Figure *2.4B***) are a smeared manifestation of these peaks.

The general pattern of a speeded re-discovery of REG and a persistent lower sustained response in the INT conditions matches the predictions of the "context incorporating" family of models. This is because the reset model does not predict faster re-discovery of REG, and the context incorporating models maintain a memory of the INT tones that directly affect the representation of the REG sequence following its resumption.

The mean amplitude patterns across INT conditions (**Figure *2.4D***, left) were consistent with an effect of INT duration on the sustained response, although there was no statistically significant difference between the INT conditions ($F(2,54) = 1.81$, $p = 0.17$; repeated-measure ANOVA). The difference from the control (INT0) appeared graded, as reflected in the pattern of significance (horizontal lines) in **Figure *2.4B***. Direct condition comparisons revealed only a small effect between INT1 and INT5 (**Figure *2.4D***, right). This is perhaps not surprising given that the conditions only differed by the introduction of 2 tones. But overall, the pattern of EEG data appears consistent with a model that maintains a long enough pre-training window to incorporate a memory of the preceding REG and the INT tones into the inferred predictability of the post-interruption REG.

Overall, these results indicate that the presence of interrupting tones affected the representation of REG even a second or more after the interruption had ended. This pattern aligns with the predictions of *context incorporating* models (Model 1) which suggest that memory of the INT tones influences the IC of the REG pattern in a manner reflected in the EEG data.

2.5 General discussion

The analysis used in this study focused on the dynamics of the EEG sustained response. Accumulating evidence suggests that it reflects the process of predictability tracking in statistically structured sequences (Barascud et al., 2016; Bianco et al., 2025; Hu et al., 2024; Zhao et al., 2025), supported by the coordinated processing of information across a distributed neural network. Source localization of the MEG sustained response (Barascud et al., 2016; Bianco et al., 2025; Hu et al., 2024) implicates a distributed network involving the auditory cortex (AC), hippocampus (HC), and inferior frontal gyrus (IFG) in representing REG patterns. This activity fluctuates dynamically, decreasing during REG interruptions and reinstating upon the discovery of a new REG pattern. These fluctuations likely reflect the disruption of top-down connectivity when an existing model is deemed no longer relevant and the strengthening of top-down connectivity when predictive models are available.

This study investigated whether and how the passive-listening brain utilises past experiences to represent ongoing sound sequences by recording EEG sustained responses in two situations: one in which a REG sequence is replaced by a different REG sequence (Experiment 1) — and another in which a REG sequence is occasionally disrupted by a varying number of new tones (Experiment 2).

2.5.1 Sustained responses to REG patterns are affected by brief interruptions

Experiment 2 revealed that sustained responses to the post-interruption REG patterns were affected by the INT tones. This finding suggests that the brain represents the post-INT REG sequence using past information, including the history of INT tones.

Prior research similarly indicates that the brain incorporates long-term sensory history when processing sequences (Maheu et al., 2019; Rubin et al., 2016; Ulanovsky et al., 2004; see also Demarchi et al., 2019; Fritsche et al., 2022), though estimates of this duration vary depending on the specifics of the

paradigm used. For instance, Rubin et al. (2016) found that auditory cortex neurons in anesthetised cats best fit prediction models accounting for more than ten previous tones (~9 seconds). Similarly, Benjamin et al. (2024) showed that tone information remains decodable from MEG responses for approximately eight successive items (2 seconds) during passive listening. Skerritt-Davis and Elhilali (2018) revealed that memory span, estimated by fitting a Bayesian perceptual model to behavioural data, correlated with performance, extending up to the full duration of each sequence (60 tones; ~19 seconds). Zhao et al. (2025) applied a similar model to random tone-pip sequences and found that most listeners based their judgments on a context of 20–40 tones (~1-2 seconds).

In the current study, the result shows that even brief contextual perturbations—such as a single interrupting tone—can alter the brain's representation of an ongoing pattern. Given the link between the sustained response and perceived predictability, the reduced sustained response amplitude following interrupting (INT) tones indicates that the inferred predictability of the REG pattern was diminished after the interruption, despite no change in the stimulus itself.

This phenomenon—where transient surprise alters neural responses to an otherwise unchanged stimulus—has been observed across multiple research domains. In post-traumatic stress disorder (PTSD), for example, neural and physiological responses to a stimulus can change after a surprising or stressful event coincides with it, even if the stimulus itself remains the same (Kaczkurkin et al., 2017; Nutt & Malizia, 2004; Sartory et al., 2013; Wessa & Flor, 2007). Similarly, in perceptual decision-making, when participants predict an image's location or value based on previous patterns, a surprising rule deviation can significantly alter their representation of the stimulus and its environment (Kao et al., 2020; McGuire et al., 2014; Nassar et al., 2010, 2012). This consistency across different psychological domains suggests a fundamental heuristic employed by the brain to track environmental changes.

2.5.2 Sustained response dynamics reflect memory of INT and pre-interruption REG

As discussed above, the sustained response modulations are consistent with a memory trace for the interruption, which causes the sustained response to settle below the level observed in the control condition (INT0). As seen also in the model, the same memory effects also influence the speed of REG re-discovery after INT. Specifically, re-discovery occurs more rapidly than the initial discovery of a new regularity (see also Bianco et al., 2025). In all cases, the sustained response begins to rise before a full cycle of the REG pattern has elapsed. Thus, both the modulation of the sustained response and the accelerated re-discovery of REG following INT reflect the system's use of prior information, aligning with the notion of a "context-incorporating" memory process, even within a simplified model framework.

Interestingly, the response dynamics also suggest memory decay. According to the model, under perfect memory conditions, the latency of REG re-discovery should be identical across INT conditions once the duration of the interruption is accounted for. However, this is not what was observe in the EEG data: following INT5, the re-discovery is delayed by approximately one tone (50 ms) compared to shorter interruptions. This delay cannot be explained by increased memory interference, as each INT condition uses a distinct set of tones, eliminating overlap. Instead, the delay points to a reduction in memory duration—i.e., decay—rather than a loss of memory content. This finding is significant because it demonstrates that memory decay can be detected within this paradigm, even during passive listening.

2.5.3 Distinct sustained response patterns in Experiment 1 and Experiment 2 suggest listeners can use or ignore context depending on its relevance

Experiment 1 yielded a different pattern of results to Experiment 2, despite the use of similar sound stimuli and the same analysis protocol. While acknowledging that Experiments 1 and 2 were conducted separately and differ

in several respects—making direct comparisons necessarily speculative—this divergence suggests that the passive-listening brain may employ a flexible context integration strategy.

In Experiment 1, the sustained response indicated that the REGy representation remained unaffected by the REGx context, as reflected in the full recovery of REGy amplitude to the REGx level (**Figure *2.1***). This pattern aligns with models where the REGx memory was either diluted by other contextual memories or erased entirely. In the simple modelling world used in this experiment, the results of Experiment 1 are either consistent with the *reset* model (Model 2) or a *context incorporating* model that learns from a relatively long prior context (e.g. Model 1.4).

In contrast, as discussed above, the pattern of results in Experiment 2 is not consistent with a *reset* model, but rather with models maintaining a memory of the preceding trials. Notably, as illustrated in **Figure *2.2***, *context incorporating* models predict a greater IC deviation from the control condition (REG, INT0) in the REGxREGy condition (Experiment 1) than in the INT5 condition (Experiment 2). However, the EEG data reveal the opposite pattern—the INT5 condition shows a larger deviation from the control (**Figure *2.5***). Therefore, to reconcile both experiments, a parsimonious conclusion is that different strategies are used by the brain in the two experiments: a “memory reset” strategy for Experiment 1 and a “memory incorporating” strategy for Experiment 2.

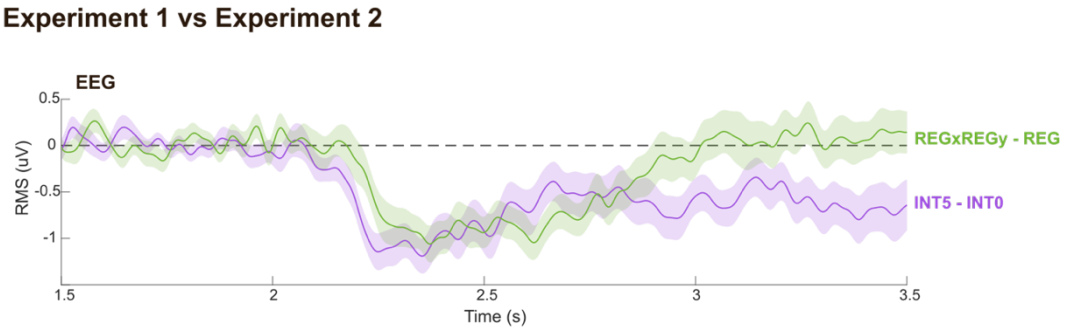


Figure 2.5 Comparing EEG across Experiments.

Group-averaged RMS of brain responses. The difference between REGxREGy and REG (Experiment 1) is shown in green. The difference between INT5 and INT0 (Experiment 2) is shown in purple. Data are baseline-corrected using the pre-transition window (1.5–2 s). Shaded areas indicate ± 2 SEM.

One possible explanation for the existence of different strategies in the two experiments lies in the distinct differences in stimulus set, and hence listeners' belief about the environment imposed in the two experiments. In Experiment 1, a violation of REGx (the first tone violating the REGx pattern) was always associated with a transition to a new pattern (REGy), meaning that REGx was not relevant, and its representation could be discarded to facilitate the learning of REGy. In contrast, in Experiment 2, the regular pattern consistently re-emerged shortly after an interruption, reinforcing the expectation that the pre-interruption REG sequence remained relevant. Therefore, in Experiment 1, participants may have automatically adapted by discarding the REGx representation, similar to the behaviour of the *reset* model, which erases prior context when making new predictions. Conversely, in Experiment 2, participants may have learned that the pre-interruption REG pattern remained relevant, leading them to preserve its memory even after the interruption. This

difference suggests the brain's ability to flexibly adjust its memory integration strategies based on the statistical structure of the auditory environment.

Importantly, a similar effect was also observed in Bianco et al. (2025; **Figure *2.6*** reproduces their findings). In that study, the authors investigated two experimental auditory environments. One, labelled 'ENVnovel', consisted of sequences that always transitioned to a new pattern (REGxREGy, as in the current study; REGxRND; and a REGx control). The other context, 'ENVresume', presented in a separate set of blocks, included REGxRND and REGx sequences but also, crucially, a condition in which the original REGx pattern resumed after an interruption by 10 random tones (REGx-RND-REGx; Unlike in the present Experiment 2, the length of the interruption was not varied). Thus, while in ENVnovel it was possible to "discover" that once REGx was disrupted, the associated predictive model was no longer relevant (since REGx would not reappear), this was not the case in ENVresume, where REGx was reintroduced 30% of the time. The results of that experiment revealed a pattern consistent with the findings reported here: the sustained response in the REGx-RND-REGx condition was consistently lower than in its control counterpart. This suggests that the brain retains information about past regularities in memory even when they are not guaranteed to reoccur.

This flexibility in adjusting the duration of reference memory is considered a crucial feature of the brain, allowing it to maintain an accurate representation of a dynamically changing environment (Bland & Schaefer, 2012; Glaze et al., 2015; O'Reilly, 2013; Yu & Dayan, 2005). Such adjustments occur in response to environmental state changes or change points—moments when past observations become unreliable for predicting future events. When a change point occurs, minimising the influence of past memory and prioritising new evidence accumulation enables a rapid adaptation to the new environment (Glaze et al., 2015; Nassar et al., 2010; O'Reilly, 2013; Skerritt-Davis & Elhilali, 2018, 2021a). Empirical studies, mostly conducted using tasks involving slow decision-making and active attention allocation, suggest that humans can

flexibly adjust their change point assumptions based on volatility estimates (Behrens et al., 2007; Glaze et al., 2015, 2018; Nassar et al., 2010). The present results suggest that similar heuristics might be operating on a faster time scale associated with sensory processing. Further controlled studies are needed to clarify these effects. For example, comparing results from conditions (REGx-INT-REGy vs REGx-INT-REGx) presented in separate blocks versus intermixed within the same block could help determine whether strategy adjustments occur on a trial-by-trial basis, at the block level, or require prolonged exposure throughout the experiment.

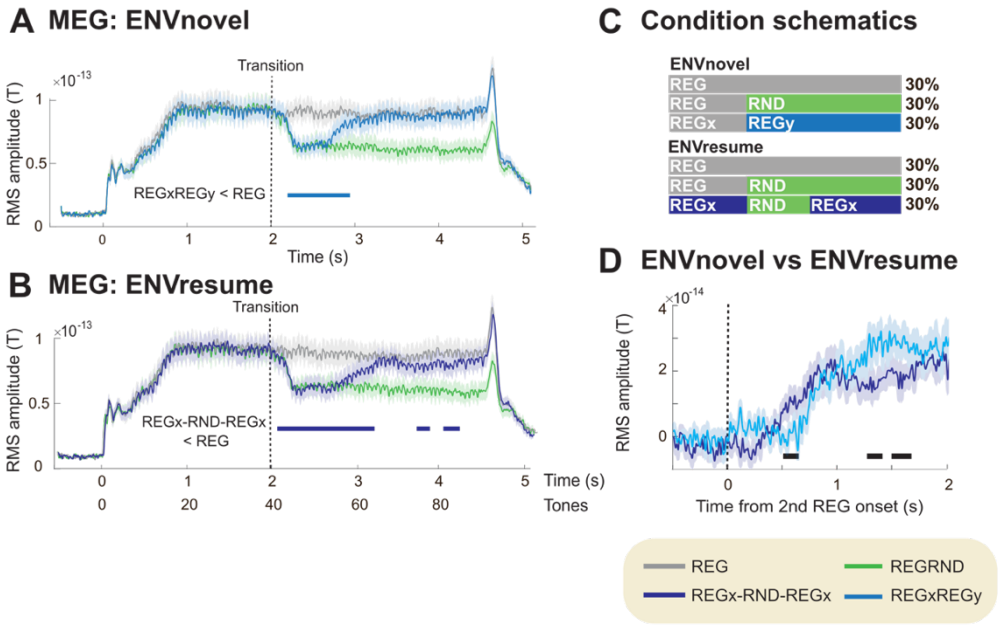


Figure 2.6 Results from Bianco et al. (2025) reveal results consistent with Experiment 2 here.

The study presented stimuli in two contexts. In 'ENVnovel', transitions were always to a new pattern (REGxREGy, as in the current study; REGRND; and a REG control). 'ENVresume', presented in a separate set of blocks, included REGRND and REG sequences but also, crucially, a condition in which the original REGx pattern resumed after an interruption by 10 random tones (REGx-RND-REGx). **[A]** Group-

average MEG brain responses (RMS across “auditory” channels; see more details in Bianco et al. (2025)) from the ENVnovel block. Data are baseline-corrected to the -0.5–0 s pre-onset window. Significant differences are indicated by the bold line below the MEG traces. These results are consistent with those in Experiment 1 here. **[B]** Data from the ENVresume block, demonstrating a persistent difference between REG and REGx-RND-REGx conditions following the resumption of REGx. **[C]** Design schematics illustrating the stimulus sequences for each condition. **[D]** Direct comparison of REGxREGy from ENVnovel and REGx-RND-REGx from ENVresume. REGRND condition data are subtracted from each condition of interest. Significant differences between conditions are indicated by bold lines below the traces. The results demonstrate a reduced sustained response in ENVresume relative to ENVnovel, consistent with the observations in Experiment 1 and 2 here.

Additionally, it is important to stress that this study use simple model comparisons focusing on varying the length of the pre-training window (consisting of counts of occurring n-grams). However, this study did not account for other model dynamics or potential parameter variations. Moreover, while the IDyOM model was used in this study due to its success in predicting human sequential processing (Barascud et al., 2016; Cheung et al., 2019; Di Liberto et al., 2020; Kern et al., 2022; Quiroga-Martinez et al., 2021), this model does not account for complex cognitive constraints, such as dynamic memory limitations and low-level auditory sensitivity. Further exploration of various models and model parameters is critical to better understand how the brain flexibly tracks the ongoing sequences under dynamic environments.

3. Chapter 3: The Effect of Prior Context Predictability on the Discovery of New Regularity

3.1 Summary

Regularity tracking is a fundamental aspect of auditory scene analysis, yet it remains unclear how this process is shaped by the statistical properties of preceding context. In this study, I examined how prior exposure to either random (RNDx) or regular (REGx) auditory sequences influences the brain's detection of a newly emerging regular pattern (REGy). Predictions were benchmarked against two computational models of statistical learning: IDyOM, a symbolic variable-order Markov model, and D-REX, a Bayesian change point detection model. However, neither model captured the sustained neural activity observed in the EEG recordings. Specifically, EEG data (N = 26; both sexes) revealed that the emergence of the neural signature for REGy was delayed when it was preceded by a deterministic (REGx) context. This finding demonstrates how prior context shapes perceptual inference and reveals critical discrepancies between computational model predictions and actual neural dynamics, suggesting the presence of brain-specific heuristics and constraints that are not yet incorporated into existing models.

3.2 Introduction

The ability to rapidly and efficiently build predictive models of the environment is fundamental for survival across species; such models allow for the optimal allocation of attentional and cognitive resources (Bendixen et al., 2012; Boubenec et al., 2017; Bouwkamp et al., 2025; Kok, Jehee, et al., 2012; Nobre et al., 2007; Southwell & Chait, 2018; Yon et al., 2018). A crucial aspect of this process is the detection of regularities in the sensory environment, which enables accurate predictions about future events (Bendixen, 2014; Bendixen et

al., 2012; de Lange et al., 2018; Friston, 2005; Press et al., 2020; Winkler et al., 2009). This is especially important in the auditory domain, where natural scenes—such as footsteps, bird chirps, or flowing water—are rich in temporal structure. Detecting and tracking these patterns supports effective monitoring of rapidly changing soundscapes (Andreou et al., 2011; Bendixen, 2014; Bendixen et al., 2012; Skerritt-Davis & Elhilali, 2018).

Previous studies have extensively investigated how the brain learns and responds to various types of regularities in auditory sequences. Notably, Barascud et al. (2016) showed that the emergence of regular structure in a sound sequence elicits a sustained neural response detectable via MEG, reflecting an online process of regularity detection. The source localisation revealed that this increase in sustained response while discovering the regularity was associated with the activation of a distributed network involving the auditory cortex, inferior frontal gyrus, and hippocampus.

However, most studies have focused on regularities emerging from silence or from acoustically random input, largely overlooking the influence of prior auditory context. In everyday listening, new regularities are rarely encountered in isolation—they are typically embedded within continuous streams of sound. Despite this, the impact of preceding auditory structure on the brain's ability to discover and track new regularities remains poorly understood. Does prior context *facilitate* the discovery of new regularities by pre-tuning the brain toward pattern detection? Does it *interfere* by anchoring predictions too strongly to outdated structure? Or might it *have no effect at all*, leaving the discovery of new regularities unaffected? These questions remain open and highlight a critical gap in our understanding of how the brain navigates dynamic auditory environments.

To investigate how prior auditory context shapes the processing of emerging regularities, I designed auditory sequences that all transitioned into the same predictable pattern (REGy) but were preceded by two distinct contexts: either another regular pattern (REGx) or a random sequence (RNDx).

Crucially, the frequency content was controlled such that REGx/RNDx and REGy were composed of distinct sets of frequencies; This ensured that any observed differences could be attributed to the statistical structure of the preceding context rather than low-level acoustic overlap (see Methods; **Figure *3.1A***). To maximise the influence of prior context on the neural representation of REGy, these conditions were presented in separate blocks. Prior works have demonstrated the brain's ability to maintain memory traces of sound streams across trials (Bianco et al., 2025; Magami et al., 2025), raising the concern that presenting REGxREGy and RNDxREGy within the same block could attenuate contextual effects. Accordingly, these responses were measured in separate blocks, alongside two additional control conditions (REG and RND), each referred as REGxREGy block and RNDxREGy block.

This design enabled an investigation into how the brain's response to the same regularity is modulated by the statistical nature of the preceding input. To interpret the neural signatures of this transition, expectations were benchmarked using two computational models of predictive processing: the Information Dynamics of Music (IDyOM) model and the Dynamic Regularity Extraction (D-REX) model. These models embody distinct approaches to evidence accumulation.

As discussed in Chapter 1 and 2, IDyOM implements a variable-order Markov model based on the Prediction by Partial Matching algorithm (Harrison et al., 2020; Pearce, 2005), and has been extensively used in studies using stimuli closely aligned with the present work (Barascud et al., 2016; Bianco et al., 2020, 2025; Harrison et al., 2020; Magami et al., 2025). As a symbolic model, IDyOM transforms frequency information into discrete tokens (i.e., alphabets) and operates in a manner that is blind to the physical properties of the tones. The model learns incrementally as the sequence unfolds and generates a conditional probability distribution for each upcoming tone based on its learned model. From this distribution, the model computes the information content (IC) of each tone —indicating how unexpected a tone is given its

preceding context. In this study, I used two variants of IDyOM. The first is a *short-term IDyOM model (STM IDyOM),* which accumulates evidence and generates predictions within each individual trial. This variant aligns with the motivation of this study to test the influence of the immediate preceding context on the processing of REGy. However, empirical evidence suggests that listeners retain statistical information beyond individual trials (Bianco et al., 2020; Kern et al., 2022; Magami et al., 2025). Motivated by this, I introduced a second variant: the *long-term IDyOM model (LTM IDyOM)*, which accumulates evidence continuously across all sequences within a block. To reflect the actual stimulus environment experienced by listeners, the model was run separately for the REGxREGy and RNDxREGy blocks.

In contrast, D-REX is based on an extension of the Bayesian Online Change Point Detection framework widely used in the sequential decision-making literature (Adams & MacKay, 2007; Nassar et al., 2010). Like IDyOM, D-REX generates a predictive distribution for the next tone based on previous observations. However, D-REX differs in several important ways. First, it assumes that input tones are sampled from a Gaussian distribution over a continuous frequency space. Second, D-REX continuously estimates the change point probability—i.e., the probability that the underlying generative distribution has changed—and updates its reference window for prediction accordingly (Skerritt-Davis & Elhilali, 2018, 2021a). When a change point probability increases, the model shortens its reference window and places greater weight on recent sensory input, allowing for rapid adaptation to the new environment. This approach makes D-REX particularly well-suited to modelling dynamic belief updating in the presence of changing statistical structure. Such adjustments of weighting between new input and prior belief is referred to as the learning rate (Nassar et al., 2010; Sutton & Barto, 1998; Williams, 1992). A higher learning rate means the model gives greater weight to the incoming input, facilitating faster adaptation to the new context by down-weighting the outdated contextual information. Empirical evidence from human decision-making studies demonstrates that learning rates adaptively increase after a rise

in change point probability, supporting the plausibility of such computations in the brain (Behrens et al., 2007; Glaze et al., 2015; McGuire et al., 2014; Nassar et al., 2010, 2012). This suggests that similar computational mechanisms may operate even during passive listening. In this study, I focused on three key statistics inferred by D-REX: (1) Surprisal: a measure analogous to IC in IDyOM. (2) Change point probability: the estimated probability that a change point has occurred at a given time. (3) Precision: the reliability of the model's predictions, quantified as the inverse of the width of the predictive distribution.

Figure *3.1B* illustrates model predictions across the STM IDyOM, LTM IDyOM, and D-REX frameworks. In the STM IDyOM, IC gradually diverges between REGx and RNDx as the REG pattern is learned. At the transition to REGy, both transition conditions exhibit a sharp spike in IC, reflecting the surprise elicited by the sudden appearance of previously unheard frequency. The spike is larger for the REGxREGy transition, as it not only introduces new frequencies but also violates an already established regular pattern. Following the transition, IC gradually decreases as the REGy pattern becomes established. The rate and pattern of IC reduction during REGy are identical for both transition conditions, as REGy frequencies were novel in both conditions.

The LTM IDyOM shows greater variability across trials (indicated by larger error bars) due to cumulative influences from prior tones. At the REGy transition, only the REGxREGy condition produces a pronounced IC spike. For the RNDxREGy condition, IC remains stable, as all 20 frequencies are nearly equally probable under the RND context, which makes the appearance of any tone less surprising. During the REGy discovery phase, the initial drop in IC, reflecting pattern learning, is similar across transition conditions. However, a small but consistent difference emerges in the sustained IC values that follows. This pattern suggests that in the REGxREGy condition, the model forms a more accurate internal representation of the REGy pattern more quickly than in the RNDxREGy condition, leading to lower ongoing IC. This advantage likely

reflects differences in the n-gram memory accumulated during the preceding context.

The D-REX model shows comparable patterns in surprisal—a measure similar to IC. Like IDyOM, D-REX generates a large spike in surprisal at the REGy transition only for the REGxREGy condition. This is because the predictive distribution in RNDx is broad, making the appearance of a new frequency unsurprising, whereas REGx's narrow predictive distribution amplifies the violation at transition.

The change probability, another key metric from D-REX, estimates the likelihood that a change point has occurred. An initial rise is observed across all conditions as the model detect the mismatch between the prior distribution and the actual input sequence. After the REGy transition, change probability rises rapidly in REGxREGy condition. In contrast, for the RNDxREGy condition, the gradual increase is seen after two full cycles of the REGy pattern. This asymmetry arises again from the fact that REGx compared to RNDx has sharper predictive distribution, making it easier to detect the emergence of the new context.

This dynamic is echoed in the precision metric, which captures the reliability of the model’s predictions. Precision has previously been shown to map closely onto the sustained neural response, a well-established proxy for regularity detection and learning (Zhao et al., 2025). Following a rise in change point probability, the model shifts its weighting towards incoming sensory input over prior expectations, thereby enhancing learning of the new structure. Indeed, in the REGxREGy condition, precision rises more rapidly following the transition compared to the RNDxREGy condition. This divergence begins around the first tone of the second REGy cycle—approximately when the change probability reaches its peak—indicating that the model has committed to the new structure. This gives REGxREGy an advantage: the clearer boundary provided by the REGx context enables the model to shift more decisively into learning the new regularity.

In summary, although IDyOM and D-REX are grounded in different computational principles, they converge on two key predictions: (1) Stronger transition-evoked responses are expected for REGxREGy than for RNDxREGy. (2) REGy is learned faster following a structured context (REGx) compared to a random one (RNDx), as captured in both LTM IDyOM and D-REX. In the following EEG experiment, I test whether neural dynamics reflect these model-derived predictions.


A Condition schematics
REGxREGy
RNDxREGy
2k
200
Frequency (Hz)
0.5
1
1.5
2
2.5
3
Time (s)
REGx / RNDx
REGy
REGxREGy block
REG
RND
REGx
REGy
RNDxREGy block
REG
RND
RNDx
REGy
B Ideal observer models
IDyOM
Short-term
IC (bits)
0
8
REG
REGxREGy
RNDxREGy
RND
Long-term
IC (bits)
0
8
Transition
10
20
30
40
50
60
70
Tones
D-REX
Surprisal
Surprisal (bits)
14
19
Change probability
Change probability
1
0
Precision
Precision
0.1
0.02
Transition
10
20
30
40
50
60
70
Tones

Figure 3.1 Stimuli and model simulations.

[A] Left: Schematic illustration of the frequency selection method, with each frequency represented as a circle. In this example, the brown frequencies were allocated to REGx/RNDx and the pink to REGy. Middle: Spectrograms depicting example stimuli for REGxREGy and RNDxREGy. The dashed line marks the onset of REGy. Right: Design schematics illustrating the stimulus sequences presented in each block type. **[B]** Model simulations. Top: Information content (IC) computed from STM and LTM IDyOM models. The y-axis is inverted (bottom = higher IC). For each condition, data are averaged over 120 trials, with shaded areas representing twice the standard error. Significant differences ($p<.01$) between REGxREGy and RNDxREGy are indicated by the grey horizontal lines. Bottom: Outputs from the D-REX model. The y-axis of surprisal is inverted (bottom = more surprising). For each condition, data are averaged over 120 trials, with shaded areas representing twice the standard error. Significant differences ($p<.01$) between REGxREGy and RNDxREGy are indicated by the grey horizontal lines.

3.3 Methods

3.3.1 Stimuli

The stimuli (**Figure *3.1A***) were 3500 ms long sequences composed of 50 ms tone pips (5 ms raised cosine ramps; 70 tone pips in total). Tone frequencies were drawn from a pool of 20 logarithmically spaced values between 222 and 2000 Hz. Successive frequencies in the pool were perceptually distinguishable from each other. The tone-pips were arranged to yield four sequence types: REG, RND, REGxREGy, and RNDxREGy. **REG**

sequences were generated by randomly selecting 10 frequencies from the pool without replacement, and this order was cycled to create a regularly repeating pattern. **RND** sequences were generated by randomly selecting 10 frequencies from the pool without replacement, and then presenting them pseudo-randomly with the following constraints: (1) The first 10 tones in the sequence had unique frequencies. (2) The same frequency was not repeated consecutively. (3) All 10 frequencies appeared equally in the sequence (7 times each). **REGxREGy** and **RNDxREGy** sequences were created by the combination of REG and RND, with the pattern transition occurring 2 s after the sound onset. To make the first part of the sequence (REGx and RNDx), the 10 frequencies were randomly selected from the pool without replacement, and the remaining 10 frequencies were used to make the second part of the sequence (REGy; **Figure *3.1A***).

The sound stimuli were organised into two block types, referred to here as REGxREGy block and RNDxREGy block. Each block contained three types of sequences: REGxREGy block included **REG**, **RND**, and **REGxREGy** stimuli, while RNDxREGy block included **REG**, **RND**, and **RNDxREGy** stimuli. To allow for direct comparisons across the transition stimuli, sequences were constructed in matched pairs. Each REG trial in REGxREGy block had a counterpart in RNDxREGy block with identical frequency content (but different pattern), and the same was true for RND trials. For the transition trials (REGxREGy and RNDxREGy), the initial segments (REGx and RNDx) were also generated in matched pairs across blocks. Ideally, the terminal segment REGy would have been kept identical across both contexts. However, due to concerns that participants might recognise and recall repeated REGy patterns (as observed in Bianco et al., 2020; Bianco et al., 2025), different REGy sequences were used in each block. These REGy sequences contained the same set of tone pips but arranged in a different order, thus maintaining the same spectral content on average while avoiding exact pattern repetition. A unique stimulus set was generated for each participant. Block order was

counterbalanced, and within each block, stimuli were presented in random order with a jittered inter-stimulus interval (ISI) ranging from 2.5 to 3 seconds.

3.3.2 Procedure

Participants were seated in an acoustically shielded room (IAC triple-walled sound attenuating booth). They listened to auditory stimuli while performing a decoy visual task, displayed on a computer screen located about 90 cm away. The visual task consisted of sequentially presented triplets of photographs of landscapes, and participants were instructed to press a key when they found that the first and third images matched, which occurred in 40% of trials. Feedback regarding the number of hits, misses and false alarms for the visual task was provided at the end of each block. The duration of image presentation was jittered between 2 and 5 s, and images were cross faded to avoid abrupt visual transients. The image presentation timing was not correlated with that of the auditory stimulus.

Participants completed the two stimulus blocks with a 10-minute break in between: Each block was divided into four 10 min runs. In total, 360 sound stimuli were presented in each block (120 stimuli per condition; in random order). The block order was randomised across participants. During the break, participants watched a cartoon video with audio to reset the memory from the previous block.

Sounds were presented diotically at a comfortable listening level through earphones (3A Insert Earphone, 3M) via a Fireface UC sound card (RME). Stimulus presentations were controlled with the Psychtoolbox package (Psychophysics Toolbox Version 3) in MATLAB (2019b The MathWorks, Inc.).

3.3.3 Recording and data processing

The general recording and data processing methods are described in Chapter 2 Experiment 1 (there, only the REGxREGy and REG data were used

for analyses). In summary, EEG signals recorded from 64 electrodes were down-sampled to 256 Hz, low-pass filtered at 30 Hz (two-pass, Butterworth, 5th-order) and detrended by a 1st-order polynomial. The data were divided into epochs of 6 s, from 1 s pre-stimulus onset to 1.5 s post-stimulus offset. The epochs were then baseline-corrected relative to the pre-onset interval (-0.5 s to 0 s relative to the sound onset). Outlier epochs and channels were removed by visual inspection, resulting in the removal of an average of 4.24 % of epochs and 0.9 channels per participant. De-noising source separation (DSS; De Cheveigné & Parra, 2014; De Cheveigné & Simon, 2008) analysis was then applied to maximise reproducibility across trials, and the data were re-referenced to the average of all channels.

To quantify the effects, the most auditory-responsive 10 channels were selected for each participant. The N1 component of the sound onset response was identified from the averaged data across all conditions. At the peak of the N1, the 5 channels showing the most positive activity and the 5 channels showing the most negative activity were considered to best reflect the brain's auditory-related activity. In the figures below, the instantaneous power of the brain response is quantified by computing the RMS (root mean square) across these channels, following a similar approach in other works (Barascud et al., 2016; Bianco et al., 2025; Magami et al., 2025; Southwell et al., 2017; Zhao et al., 2025). The RMS reflects instantaneous power of the brain response irrespective of its polarity. Field maps at relevant time points are also provided.

To uncover activity potentially masked by the slow DC changes, the same analysis was applied to high-pass filtered data at 2 Hz (two-pass, Butterworth, 4th-order). To extract the MMN response, and to enable us to compare its dynamics across conditions, DSS, in this analysis, was applied only to RNDxREGy and REGxREGy conditions (0-4 s post-sound onset). For each participant, the first two DSS components were retained and projected back into sensor space (including to REG and RND conditions). To select the electrodes best reflecting the MMN-like response, the data across all conditions across all

participants were averaged and the 10 electrodes with the most negative activation at the typical MMN response interval (150-200 ms post-transition onset) were selected. The output data were averaged across these channels for each condition and baseline corrected just before the onset of the transition (1.8-2 s post-stimulus onset).

To investigate brain responses associated with the discovery of REGy in REGxREGy and RNDxREGy conditions, three analyses were conducted: **REGy discovery onset analysis**, **REGy discovery offset analysis**, and **REGy response slope analysis**. All used the same data (non high pass filtered) and specifically focused on the time window starting from 2.5 s post-stimulus onset (onset of the 2nd cycle of REGy) as this is the earliest time point where the REGy pattern can be detected. The **REGy discovery onset analysis** focused on the timing of divergence between REGy and the RND control. To accurately compare the divergence time, the data were baseline corrected for each condition right before the onset of the second cycle of REGy (0.2-0.5 s post transition), i.e. just before the pattern began to repeat. These baseline-corrected REGxREGy and RNDxREGy data were then compared against the RND condition from the corresponding block using bootstrap resampling and the earliest significant timepoint was interpreted as indicating the onset of REGy discovery. Repeating the boostrap analysis allowed us to generate a distribution of this onset timing. See the statistical analysis section for the details of the bootstrap resampling method. **The REGy discovery offset analysis** focused on the timing where the REGy response reached the level of the REG (no change) control. The general analysis procedure is identical to that of the REGy discovery onset analysis, but here the data were baseline corrected between 1 s to 1.3 s post-transition onset, where the REGy amplitude plateaued in all conditions. Lastly, the **REGy response slope analysis** directly compared REGxREGy and RNDxREGy conditions. To compare two conditions recorded in separate blocks, RND in the corresponding block was subtracted from

REGxREGy and RNDxREGy for each subject data. These data were then baseline corrected at 2.4-2.5 s post-stimulus onset.

To confirm if REGxREGy block and RNDxREGy block yield different patterns in their control conditions (REG and RND), DSS was applied only to the conditions of interest (REG from two blocks or RND from two blocks; 0-4 s post-sound onset). For each participant, the first three DSS components were retained and projected back into sensor space. The same 10 electrodes were used as those used in the RMS analysis.

3.3.4 Statistical analysis

To statistically evaluate the effect of the prior context predictability on the next context discovery process, the differences between sound conditions were calculated for each participant. This difference was then subjected to bootstrap resampling (10000 iterations; Efron & Tibshirani, 1994). The difference between conditions was considered significant if the proportion of bootstrap iterations falling above or below zero exceeded 95% ($p<.05$) for more than 8 adjacent samples (Barascud et al. 2016) in all analyses. For reference, the result with the threshold of $p<.01$ was also reported. This statistical analysis method was used in the RMS analysis, MMN analysis, and REGy response slope analysis.

For the REGy discovery onset analysis, the same bootstrap resampling procedure described above was applied, comparing either REGxREGy or RNDxREGy against RND across 1,000 iterations. In each iteration, the first time point after 0.5 seconds post-transition at which a significant difference ($p<.01$) emerged was recorded. A similar procedure was used for the REGy discovery offset analysis. Here, each condition was compared to the REG baseline and the last time point before 1 second post-transition at which significance ($p<.01$) was observed was recorded.

To examine block effects on the control conditions, sustained response amplitudes were compared using the bootstrap resampling procedure described above. In addition, I assessed whether the variability of responses differed

between blocks by evaluating differences in standard error. Specifically, 1,000 bootstrap resamples of participants were performed and the standard error of the sustained response across participants for each iteration was computed. The resulting distribution of 1,000 standard errors was then subjected to further bootstrap resampling to test for significant differences between block types.

3.3.5 Modelling

In the introduction, I presented two computational models: IDyOM and D-REX. The IDyOM model was implemented using the new_ppm_simple function from the ppm R package (available at https://github.com/pmcharrison/ppm). All parameters were set to their default values as specified in the repository documentation, with one exception: update exclusion was disabled to allow a fair comparison between transition conditions. Two variants of IDyOM were used: a **short-term model (STM)**, in which the model was reset after each trial, and a **long-term model (LTM)**, in which the model accumulated information continuously within each experimental block (either a REGxREGy or RNDxREGy block).

The D-REX model was implemented in MATLAB using the run_DREX_model function (available at https://github.com/JHU-LCAP/DREX-model). All parameters were again set to their default values according to the repository documentation. The temporal dependence parameter D was set to 10—the same as the number of unique tones in one cycle of the REG sequence—to ensure that the model could reliably detect the embedded regularities. This parameter determines how many successive observations are treated as statistically dependent within the input sequence. The model was reset after each trial.

3.3.6 Participants

Twenty-eight paid participants participated in this experiment. All reported no history of hearing or neurological disorders. Two participants were excluded due to exceptionally noisy EEG data. Data from the remaining twenty-six

participants (19 females; average age 24.81, ± 4.20) were used for analyses. All experimental procedures were approved by the research ethics committee of University College London, and written informed consent was obtained from each participant.

3.4 Results

3.4.1 The response to the emergence of REGy differs depending on the preceding context

Both the REGxREGy and RNDxREGy conditions involve a *context change* triggered by the introduction of new frequencies. I first examined how the brain responds to these novel events as a function of the preceding context. Because REGx and RNDx are matched in their frequency content (see Methods), any observed differences in neural responses can be attributed to the predictability of the pre-transition sequences, i.e., regular versus random structure, rather than to spectral differences.

To isolate the MMN-like response, which I expected to be evoked at the transitions (by the first REGy tone, which constitutes a novel frequency), the EEG data were high-pass filtered at 2 Hz to eliminate any sustained differences between conditions (see **Figure *3.3***). The data were then averaged across trials for each condition. Bootstrap resampling revealed a significant deflection in the REGxREGy condition relative to its baseline control (REG, **Figure *3.2A***), with differences emerging at ~130 ms post-transition ($p<.05$). This latency, and the corresponding topography (**Figure *3.2A***) are consistent with those commonly associated with the MMN response (Winkler, 2007). Notably, similar activity was not observed in the RNDxREGy condition relative to its control condition (RND, **Figure *3.2A***).

To directly compare the MMN response between REGxREGy and RNDxREGy, the control conditions (REG for REGxREGy and RND for

RNDxREGy) were subtracted from the transition conditions (**Figure *3.2B***). Bootstrap resampling revealed a significantly larger MMN response in REGxREGy than RNDxREGy, indicating that the predictability of the preceding context affected the MMN-linked salience of a deviant tone. This finding aligns with computational model predictions: both IDyOM and D-REX estimated a sharper transition-related spike in IC and surprisal, respectively, in the REGxREGy compared to RNDxREGy condition (**Figure *3.1B***). As both IC and surprisal reflect the unexpectedness of a tone given its prior context, the MMN—an established neural index of deviance detection—is likely linked to these peaks in model-derived surprise.

Following this negativity response, the positive response was observed in the REGxREGy condition relative to the control condition (REG), where the significance emerged at ~240 ms post-transition ($p<.05$). This latency and the corresponding topography are consistent with P3a response, which often shows the front-central activity at around 200-350 ms after the onset of the surprising event (Bendixen 2007). A similar, but delayed, response was observed in the RNDxREGy condition, emerging around 290 ms post-transition ($p < .05$). Notably, in both conditions, this positivity preceded the emergence of the REGy pattern (which occurs 500 ms, or 10 tones, after the transition), suggesting that this response reflects the brain's detection of a deviation in frequency content from the preceding auditory context, not the discovery of REGy.

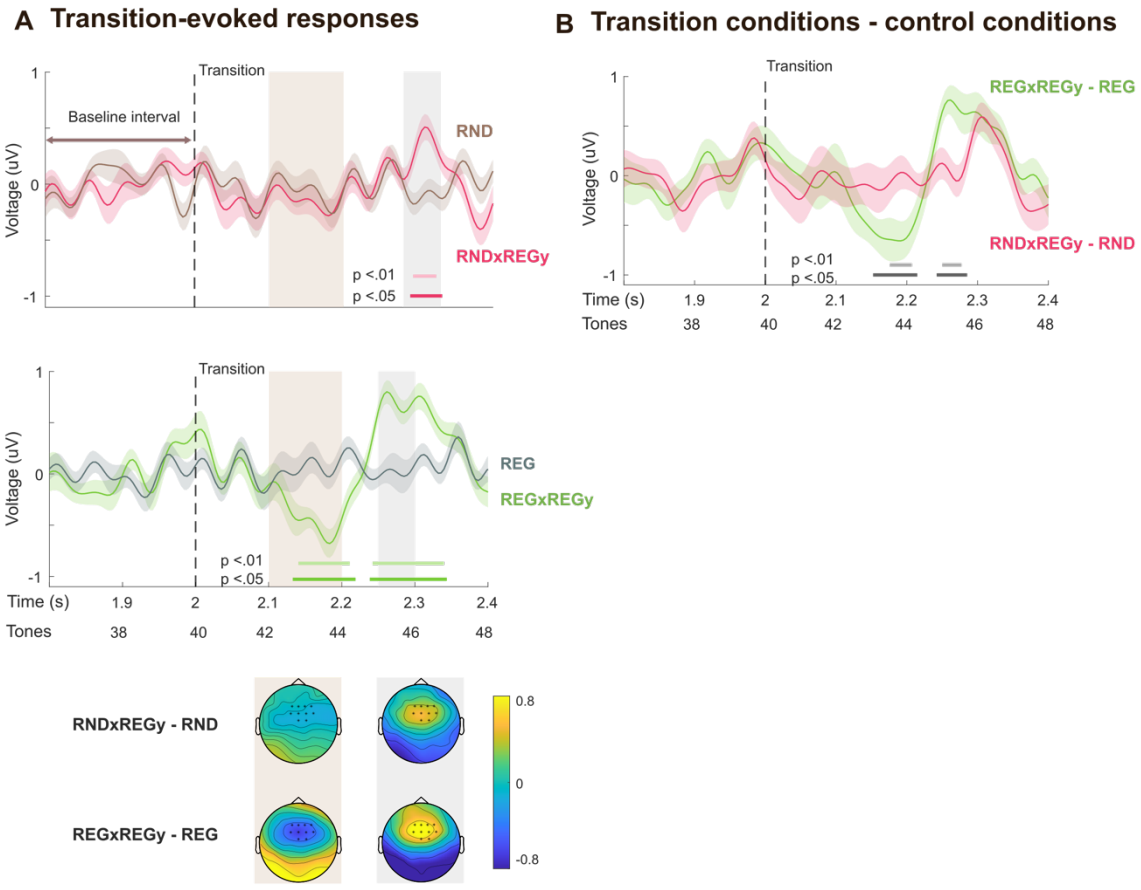


Figure 3.2 Transition evoked deviance responses.

[A] High-pass filtered mean EEG data, averaged over 10 channels (indicated in the scalp topographies). Shaded areas represent twice the SEM, computed with bootstrap resampling (1000 iterations). Significant differences (p<.01 and p<.05) between conditions are indicated by the horizontal lines below the EEG traces. Top: RNDxREGy vs RND conditions. Shaded boxes indicate the time windows (2.1-2.2 s and 2.28-2.33 s) used for scalp topography calculations plotted at the bottom of the figure. Bottom: REGxREGy vs REG conditions. Shaded boxes are located at 2.1-2.2 s and 2.25-2.3 s. **[B]** Control conditions (REG, RND) are subtracted from transition conditions (REGxREGy, RNDxREGy).

3.4.2 The EEG sustained response tracks regularity discovery and violation

For each participant and condition, the RMS over 10 selected channels (detailed in the Methods section) was calculated at each time point on the non-high-pass filtered data. The group averaged responses for two blocks (REGxREGy block, RNDxREGy block) are shown in **Figure *3.3***. For the REG and RND conditions, the brain indicated an N1 peak at around 100 ms post-stimulus onset, followed by increase in amplitude until it reached to a plateau. The timing and amplitude each condition reached a plateau differed: the condition difference between REG and RAN emerged before the end of the 2nd cycle (**Figure *3.3***). This timing aligns with previous literature (Barascud et al., 2016; Bianco et al., 2025; Magami et al., 2025) and suggests that the brain discovers the regular patter in less than 2 cycles of exposure.

For the REGxREGy condition, the amplitude abruptly dropped following the emergence of the REGy pattern. Upon the discovery of the new pattern, this amplitude gradually recovered to the pre-transition level. Bootstrap resampling revealed a significant difference between REG and REGxREGy, starting from 220 ms (~ 4 tones) after the transition ($p<.05$), consistent with previous literature (Barascud et al., 2016; Bianco et al., 2025; Magami et al., 2025). Bianco et al. (2025) attributed this drop to reduced activity in the frontal-auditory network, suggesting that the brain's predictive model of the REGx pattern was disrupted. For the timing of REGy discovery, the responses between the REGxREGy and RND conditions were compared to determine when REGy began to diverge from randomness. A significant difference between RND and REGxREGy emerged at 885 ms post-transition (~ 1 cycle + 8 tones).

Lastly, the RNDxREGy condition did not show an abrupt response to the transition. Rather, it indicated a gradual increase in amplitude as REGy emerged. Bootstrap resampling indicated that the divergence from RND

condition started from 770 ms (~1 cycle + 5 tones). This is roughly 100 ms faster than REGxREGy condition.

These dynamics were well captured by both the IDyOM and D-REX models (**Figure *3.1B*).** For IDyOM, initial discovery of the REG pattern occurred during the second cycle, closely aligning with the EEG data. The model's response trajectories across four conditions (REG, RND, REGxREGy, and RNDxREGy) matched the neural responses, and notably, the LTM version successfully captured the EEG-observed discrepancy between REGxREGy and RNDxREGy during the transition to REGy. For D-REX, I focused specifically on the Precision parameter, consistent with Zhao et al. (2025). In this model, REG was initially discovered by the end of the second cycle, broadly consistent with EEG, though slightly slower than IDyOM predictions. This lag likely reflects the fact that D-REX begins with a generic default distribution and incrementally updates it as new input is received. Importantly, D-REX also captured the contrast between REGxREGy and RNDxREGy conditions at transition point, consistent with the EEG observations.

Despite these broad correspondences, several key differences emerged between the models and the neural responses. First, while both models continued to refine its predictability representation throughout the REG sequence (evidenced by a steady decline in IC and increase in precision), the EEG response plateaued after an initial build-up. Second, the EEG showed a delayed response to the REGxREGy transition, with the drop in sustained activity occurring roughly four tones later than the spike responses in both models. Third, in the models, REGy was discovered either equally or more rapidly in the REGxREGy condition than in the RNDxREGy condition (see Section 3.2 for detailed explanations). Contrary to this prediction, however, the EEG data revealed the opposite trend: the divergence from the RND condition was slower in the REGxREGy condition. In the next section, I focus specifically on the pattern of REGy discovery and investigate this discrepancy in more detail.

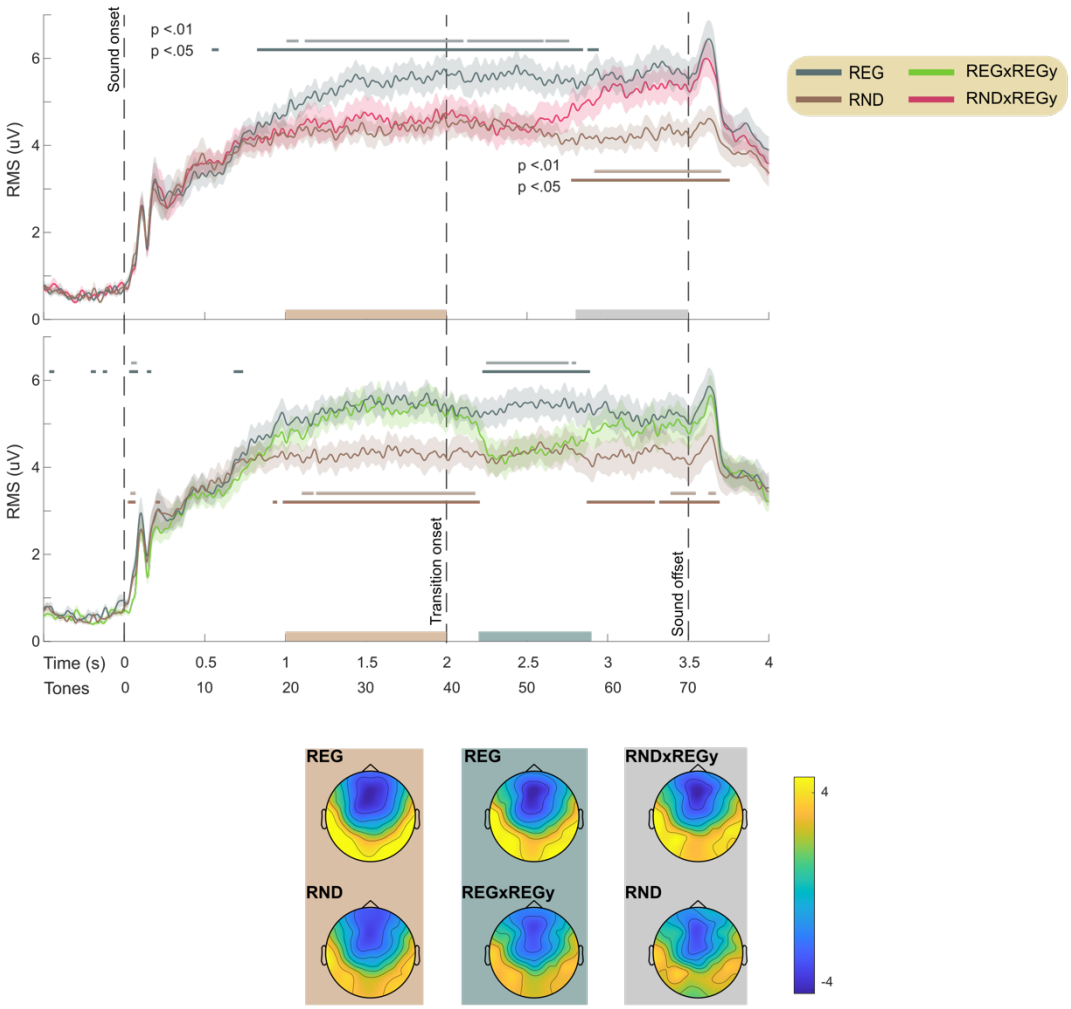


Figure 3.3 Sustained response dynamics.

Group-averaged RMS of brain responses. Shaded areas represent twice the SEM. Data are baseline-corrected to the -0.5-0 s pre-onset window. Significant differences (p<.01 and p<.05) between REG and transition conditions are indicated by grey bold horizontal lines above the EEG traces. Significant differences between RND and transition conditions are indicated by brown bold horizontal lines below the EEG traces. Top: Responses from the RNDxREGy block. Middle:

Responses from the REGxREGy block. Bottom: Scalp topographies are provided for three time windows: 1-2 s (REG from REGxREGy block and RND from RNDxREGy block), 2.2-2.9 s (REG and REGxREGy from REGxREGy block), and 2.8-3.5 s (RND and RNDxREGy from RNDxREGy block).

3.4.3 The prior context influences the discovery process of the following regularity pattern

Here, the REGy discovery pattern was compared across the two transition conditions (REGxREGy, RNDxREGy) using three metrics: REGy discovery onset timing, REGy discovery offset timing, and REGy response slope pattern.

Discovery onset was defined as the point at which the REGy response significantly diverged from the RND response. To assess this, the RMS data were baseline corrected (2.2-2.5 s relative to the stimulus onset) and bootstrap resampling was used to estimate the earliest significance divergence point (see Methods). Repeating this analysis yielded distributions of discovery onset timings, revealing a clear difference between the two transition conditions (**Figure *3.4A right***). Specifically, the discovery was *earlier* following RNDx than following REGx.

Discovery offset was defined as the point at which the REGy response aligned with the REG condition. RMS data were baseline-corrected in the post-discovery time window (3–3.3 s post-stimulus onset) and the time at which the transition condition responses converged with the REG response was identified (**Figure *3.4B***). Bootstrap resampling showed that offset timing was also later for REGxREGy than RNDxREGy.

Finally, I directly compared the discovery slope between the two conditions. Because REGxREGy and RNDxREGy were presented in separate

blocks, the RND baseline was subtracted from each block to isolate the REGy-related change. Bootstrap resampling (see Methods) showed that the rise in amplitude occurred earlier in the RNDxREGy condition, indicating a faster adaptation to REGy (**Figure *3.4C***). However, this difference reached significance only with p threshold of 0.05, but not 0.01, suggesting the effect is relatively weaker than other findings.

Taken together, these results consistently show that prior context influences the processing of emerging regularities; presence of prior REG patter delays the discovery of the new regularity. Notably, this pattern diverges from the models' predictions, which indicated the opposite.

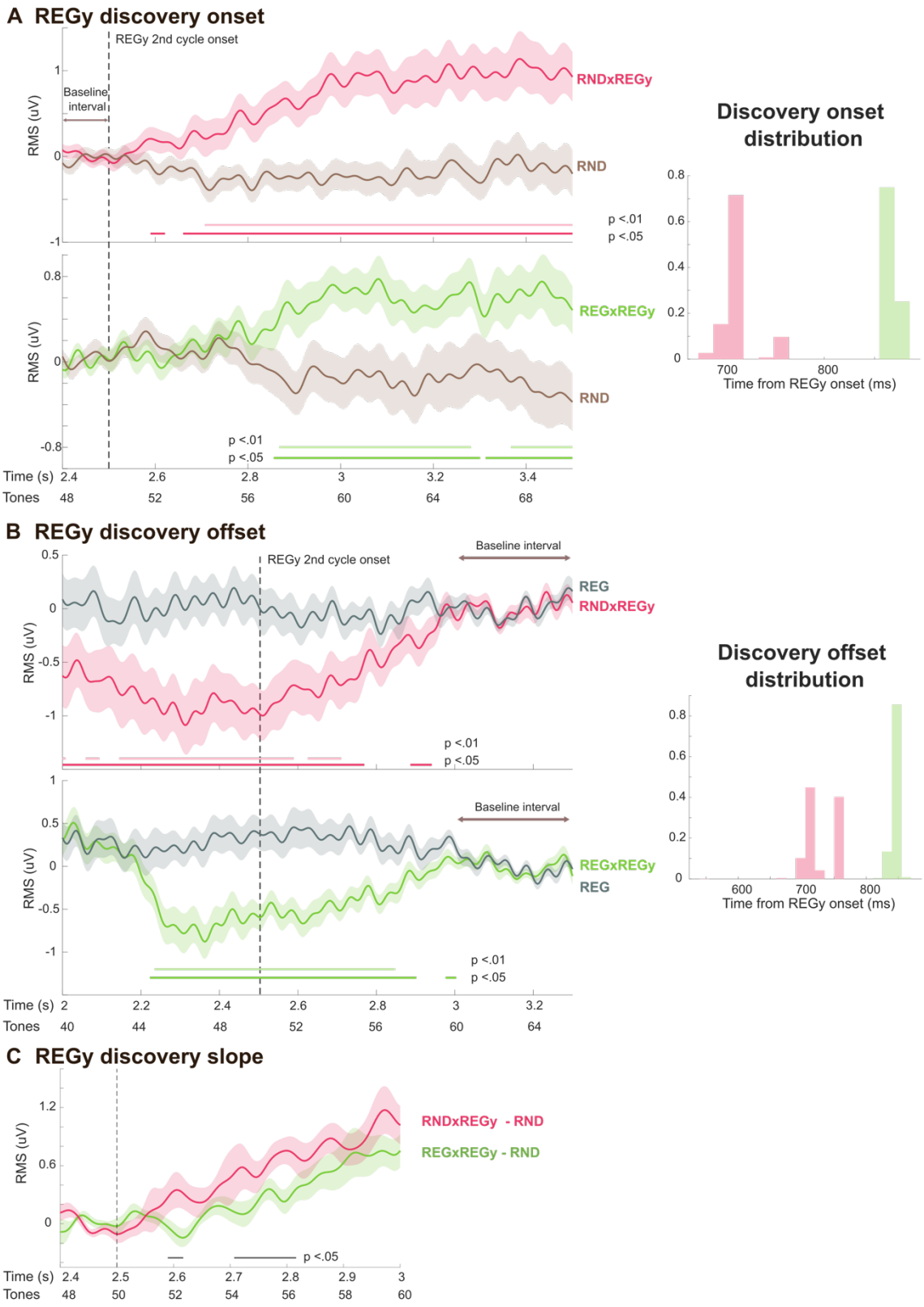

A REGy discovery onset
REGy 2nd cycle onset
Baseline interval
RMS (uV)
RNDxREGy
RND
p <.01
p <.05
REGxREGy
RND
Time (s)
Tones
Discovery onset distribution
Time from REGy onset (ms)
B REGy discovery offset
Baseline interval
REG
RNDxREGy
REGxREGy
C REGy discovery slope
RNDxREGy - RND
REGxREGy - RND

Figure 3.4 REGy discovery dynamics.

[A] RND and transition conditions are compared to determine the REGy discovery onset. Left: Group-averaged RMS of brain responses. Shaded areas represent twice the SEM. Data are baseline-corrected to 2.2-2.5 s. Significant differences ($p<.01$, $p<.05$) between RND and RNDxREGy (top) or REGxREGy (bottom) are indicated by bold horizontal lines below the EEG traces. Right: The histogram showing the latencies associated with REGy discovery onset of two transition conditions calculated by repeating the bootstrap resampling in [A]. Results are shown for the resampling thresholds of $p < .01$. They demonstrate faster discovery onset in the RNDxREGy condition. **[B]** REG and transition conditions are compared to determine the REGy discovery offset. Left: Same data as in [A] but baseline-corrected to 3-3.3 s. Significant differences ($p<.01$, $p<.05$) between REG and RNDxREGy (top) or REGxREGy (bottom) are indicated by bold horizontal lines below the EEG traces. Right: The histogram showing the latencies associated with REGy discovery offset of two transition conditions. Results are shown for the bootstrap resampling thresholds of $p < .01$. They demonstrate faster discovery offset in the RNDxREGy condition. **[C]** EEG data (RMS) for each of the transition conditions after subtracting the RND condition, baseline-corrected to 2.4-2.5 s. Grey lines beneath the traces mark significant differences ($p<.05$) between conditions.

3.4.4 Transition conditions do not yield a block-level effect

Thus far, the focus has been on how the immediate preceding context influences the discovery of REGy. However, this design also introduces an asymmetry in the global context across blocks—specifically, REGxREGy blocks contain more instances of regularity (REGx) compared to RNDxREGy blocks. To assess whether this broader contextual difference influences neural

responses to control conditions, I examined the sustained responses to REG and RND sequences presented within each block type (**Figure *3.5***).

For REG, both the amplitude and variability of the response remained stable, suggesting it was unaffected by the global context. In contrast, the RND condition exhibited greater variability across participants when RND was presented within a predominantly regular block (REGxREGy block). This suggests that reduced exposure to less-structured input (RND) may increase inter-subject response variability. RND was used as a control in some post-transition analyses described above, but differences between blocks were confined to the pre-transition period, with no significant post-transition effects observed, assuring the use of these RND conditions in those analyses. Overall, these findings indicate that introducing two different transition conditions did not exert a widespread influence at the block level.

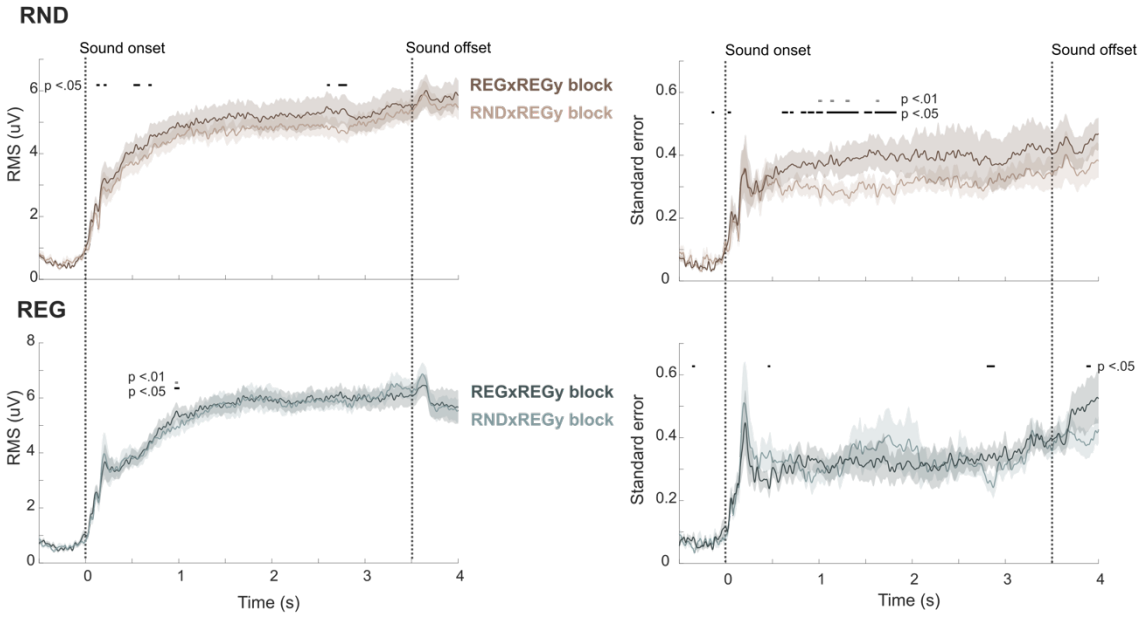


Figure 3.5 Comparisons of control conditions across blocks.

Control conditions (RND: top, REG: bottom) across two block types (REGxREGy: expressed as darker colours, RNDxREGy: expressed as lighter colours) are compared. Left: Group-averaged RMS of brain responses. Shaded areas represent twice the SEM. Data are baseline-

corrected to -0.5-0 s. Significant differences ($p<.01$, $p<.05$) between block types are indicated by bold horizontal lines above the EEG traces. Right: Bootstrap resampled the data on the left and for each iteration, calculated the standard error. The average of those iterated standard error is plotted for each condition. Shaded areas represent twice the STDEV. Significant differences ($p<.01$, $p<.05$) between block types are indicated by bold horizontal lines above the EEG traces.

3.5 Discussion

This study set out to examine how prior auditory context influences the brain's ability to detect emerging regularities, leveraging two established computational models of statistical tracking—IDyOM and D-REX. These models provided a theoretical framework for how regularities might be discovered based on the statistical properties of preceding sequences. To test these predictions empirically, I recorded sustained neural responses, a well-established marker of the brain's sensitivity to regularity. Surprisingly, the results diverged from the model predictions; the neural data revealed a delayed response when the new pattern followed a regular rather than a random sequence.

3.5.1 Deviation responses are influenced by the predictability of the prior context

In this experiment, both transition conditions (REGxREGy and RNDxREGy) had the onset of REGy marked by the introduction of entirely new frequencies not present in the preceding context. On the surface, this frequency change should be equally surprising in both cases. However, a clear mismatch negativity (MMN) response was observed only in the REGxREGy condition. The MMN is a well-established neural marker of deviance detection, and its

amplitude is considered to reflect the perceptual salience of the deviant event (Bendixen et al., 2012; Näätänen, 2001; Winkler, 2007; Winkler et al., 2009).

The role of context in shaping deviance processing has been demonstrated in various studies (Garrido et al., 2013; Herrmann et al., 2015; Khouri & Nelken, 2015; Schröger & Roeber, 2021; Southwell & Chait, 2018). For instance, Southwell and Chait (2018) employed REG and RND sequences similar to those used in the present study and occasionally introduced deviant tones. Crucially, these deviant tones were drawn from a frequency range outside the one used to construct the REG and RND contexts, making them inherently salient regardless of the background. Despite this clear physical distinctiveness, neural responses to these deviants were reduced when they occurred within a random (RND) sequence. This finding demonstrates that deviance detection is not solely driven by the physical novelty of a sound, but is also shaped by the statistical regularity of the preceding auditory context.

The present findings extend this line of research by examining transitions in which the transition (deviant) tones were drawn from the same frequency pool as the preceding context, yet never experienced in the preceding context. The observation that only the REGxREGy condition elicited an MMN response suggests two key insights: (1) when tones fall within a familiar frequency range experienced in the preceding context, their mere novelty does not strongly evoke surprise; and (2) a strong, structured prediction based on prior context can amplify the salience of a tone when it violates that prediction. In other words, even when the tone itself is not inherently surprising, it becomes surprising through the lens of violated expectations—a core idea in predictive coding frameworks (Friston, 2005). This highlights the critical role of contextual predictability in shaping the perceptual salience of auditory events.

Interestingly, both computational models predicted the observed asymmetry in MMN responses. In the IDyOM, the STM variant showed a small but distinct peak in response to the RNDxREGy transition. However, this

response disappeared in the LTM IDyOM, which integrates memory over a longer timescale. This implies that once the full frequency range is experienced and expressed in the n-grams, the appearance of a “new” tone (in terms of local context) is no longer surprising. Reflecting on this, the brain’s response may not solely be based on local transitions but also incorporate prior exposure over longer timescales than a single trial. Indeed, prior research suggests that memory for tone statistics can span several seconds to tens of seconds, depending on task structure (Benjamin et al., 2024; Rubin et al., 2016; Skerritt-Davis & Elhilali, 2018; Zhao et al., 2025), demonstrating the brain’s capacity for sustained statistical learning.

The D-REX model shows a similar pattern but for a different computational reason. In D-REX, predictions are made over a continuous frequency space. A regular context yields a narrower predictive distribution than a random one, making transition tones more likely to be flagged as violations. This differential response aligns with empirical findings by Garrido et al. (2013). They presented tones drawn from Gaussian distributions with the same mean (500 Hz) but different variances (standard deviation of either 0.5 or 1.5 octaves). Deviant tones—identical across conditions and set two octaves above the centre frequency—elicited stronger MMN responses when embedded in the narrower distribution. Furthermore , Schröger and Roeber (2021) found that when the underlying sound sequence was stochastic, deviant (i.e., rare) tones evoked an MMN response only when they fell outside the distribution of the standard (i.e., common) tones. In contrast, rare tones that remained within the distribution did not elicit an MMN. This result echoes the present findings: even when a tone is rare—or novel, in my case—it fails to evoke an MMN if it falls within the expected distribution of a stochastic sequence.

Another notable neural signature observed at the transition point was a positive deflection following the initial negative response, occurring around 250–300 ms post-transition. The timing and scalp topography of this response are consistent with the P3a component, which is typically associated with

involuntary attentional shifts toward novel or salient stimuli (Bendixen et al., 2007; 2008). Interestingly, this P3a-like response was observed in both transition conditions, in contrast to the MMN, which was condition-specific. Traditionally, the presence of MMN in a rule-violation versus rule-confirmation contrast is interpreted as evidence that the brain has extracted the underlying regularity (Näätänen & Winkler, 1999). It would be intuitive to assume that a rule-violating sound becomes salient and therefore elicits a P3a. However, how can we account for a P3a response in the absence of an MMN?

Coy et al. (2024) explored this issue using a modified oddball paradigm where standard tones were occasionally replaced by deviants. In one condition, deviant tones tended to repeat, while in another they were typically followed by a return to the standard. They compared responses to post-deviant standard tones that were either expected or unexpected. This manipulation failed to elicit an MMN but did produce a P3a-like response. Behavioural results further supported the notion that people could extract such rules, implying that the absence of MMN should not be taken as evidence against rule learning.

These findings, along with the present results, are consistent with the idea that deviance responses are hierarchically organised, as demonstrated in the local-global paradigm. In a study by Wacongne et al. (2011), participants passively listened to five-tone sequences (e.g., AAAAB–AAAAB...). A local deviant (the final B) violated the immediate tone pattern and elicited MMN responses. In contrast, a global deviant (a final A in an otherwise AAAAB context) preserved local regularity but violated a higher-order, global pattern, eliciting later P3 responses. These results suggest that the MMN reflects lower-level sensory prediction errors, whereas the P3 is associated with higher-level or context-based violations.

In the present experiment, a similar hierarchical interpretation may apply. In the RNDxREGy condition, a shift in the frequency content of the auditory scene may have triggered a P3a response, even though no clear MMN was

observed. At the latency of the P3a, the REGy sequence had not yet completed a full cycle, making it unlikely that the underlying regularity had been detected. Thus, the only available cue distinguishing REGy from the preceding random sequence was the change in frequency composition.

However, the interpretation of P3a in the absence of MMN, and more broadly, the nature of P3a itself, remains a topic of ongoing debate (Coy et al., 2024; Dien et al., 2004). Further research is needed to clarify which specific aspect of the RND-to-REG transition drives this response.

3.5.2 REGy discovery dynamics differ between EEG responses and model predictions

The sustained neural response to REGy revealed a clear influence of the predictability of the preceding context. When REGy followed a random sequence, the new regularity was processed more quickly than when it followed another regular sequence (**Figure *3.4***). Interestingly, this pattern runs counter to the predictions of the two benchmark computational models. This discrepancy is surprising, given that both models have successfully accounted for a wide range of human auditory behaviours. The IDyOM model, for example, has been shown to capture aspects of regularity discovery in both passive (Barascud et al., 2016; Bianco et al., 2025; Magami et al., 2025) and active listening (Barascud et al., 2016; Bianco et al., 2020), as well as cortical responses during music perception (Di Liberto et al., 2020; Kern et al., 2022) and subjective experiences of musical pleasure (Cheung et al., 2019). Likewise, D-REX, based on a change-point detection framework, is specifically designed to handle sequences with shifts in underlying statistics and has been shown to successfully predict behavioural change detection performance (Skerritt-Davis & Elhilali, 2018). Furthermore, Zhao et al. (2025) used stochastic tone sequences that transitioned from a broad to a narrower distribution (variants of RNDREG). The gradual increase in sustained neural activity observed as the distribution changed was closely mirrored by the rise in precision within the D-

REX model, underscoring its utility in modelling the dynamics of statistical learning.

The mismatch between these model predictions and the observed EEG results suggests that a distinct neural process, that is not accounted for in modelling, may be engaged during the passive detection of abrupt statistical transitions. In the following sections, I explore possible explanations for this discrepancy in more detail.

3.5.2.1 Biological constraints as a potential source of discrepancy

According to the predictive coding theory, the brain builds internal models based on past experience and continuously updates them by minimising prediction errors (Friston, 2005, 2008; Rao & Ballard, 1999). However, in environments where the underlying statistical structure changes, relying on the full history of past inputs may impede effective learning, as earlier information may reflect an outdated structure. Detecting such change points and adjusting the temporal window of reference is therefore crucial for maintaining an accurate internal model of the auditory environment (Nassar et al., 2010; Skerritt-Davis & Elhilali, 2021b).

Yet, identifying true change points is inherently difficult. From the observer's perspective, it is often unclear whether a given prediction error arises from random noise or an actual shift in the environment (Nassar et al., 2010; Piray & Daw, 2024; Skerritt-Davis & Elhilali, 2021b). One potential solution is to maintain multiple hypotheses about possible change points and weigh them according to their likelihood—an approach used in the full Bayesian inference models (Adams & MacKay, 2007; Nassar et al., 2010; Skerritt-Davis & Elhilali, 2018), including D-REX.

If the brain follows a similar procedure to the D-REX model, why then did the observed neural dynamics diverge from its predictions? One plausible explanation is that, although the brain may implement a full Bayesian updating,

its biological constraints impose temporal delays not present in the model. While the D-REX model can adjust its weighting immediately after detecting a change point—reflected in the very next prediction—the brain likely undergoes a sequence of biological operations: detecting the change, modulating the relative influence of bottom-up versus top-down signals, and resetting prior beliefs. These steps may unfold over a longer timescale. Thus, even if change detection occurs efficiently at the REGxREGy transition, the implementation of belief updating in the brain may not be instantaneous, potentially obscuring any speed advantage observed in the model.

3.5.2.2 The brain may not implement full-Bayesian updating

Another possibility is that the brain does not implement full Bayesian updating. Maintaining and updating predictive distributions over all possible change point hypotheses is computationally demanding. A growing body of research suggests that the brain may not always operate according to full Bayesian principles. In many situations, simpler heuristic strategies offer equally good—or even superior—accounts of behaviour (Nassar et al., 2010, 2012; Payzan-LeNestour & Bossaerts, 2011). For instance, Nassar et al. (2010) showed that a simplified model, which considers only two possibilities on each trial—either the observation originates from the same distribution or from a new one—can capture human behaviour as effectively as a full Bayesian model. However, such heuristic approaches involve trade-offs. By not tracking every possible change point, the brain increases its risk of missing true changes in the environment. This raises a key question: how does the brain maximise sensitivity to genuine changes while minimising the risk of misinterpreting noise as change?

One plausible strategy is to impose a brief “wait-and-see” period following the detection of a violation. During this window, the brain may accumulate further evidence before committing to resetting its internal model. Such a strategy would help prevent unnecessary resets triggered by noise, but

at the cost of requiring additional observations to confirm a change—potentially explaining the delay in discovering the new REGy pattern compared to the model.

Indeed, the EEG data in the current experiment indicated a delay of approximately four tones—relative to the model—before a sharp drop in the sustained response emerged following a violation of the original REG pattern (**Figure *3.3***). This observation aligns with previous findings (Barascud et al., 2016; Bianco et al., 2025; Magami et al., 2025), yet it is not accounted for by either model, suggesting that it reflects a distinctive feature of the brain's computational architecture.

Importantly, this delay is unlikely to reflect a fixed, circuit-level lag in encoding violations. Bianco et al. (2025) showed that the number of tones required for the drop remained constant even when tone durations were halved, indicating that the delay is tied to information content rather than absolute time. Further support comes from Chapter 2, where I introduced a one-tone interruption within a REG sequence—shorter than the typical "wait-and-see" window—and observed a less pronounced drop in sustained neural activity compared to longer interruptions (***Figure 2.4***). Together, these findings suggest that the delay serves as a strategic evidence accumulation period, during which the brain evaluates whether the experienced prediction violation is a spurious event or indicative of a consistent stimulus change. This delay may underlie the discrepancies observed between neural data and both models.

Taken together, neither D-REX nor IDyOM fully captured the delayed regularity discovery process observed in the REGxREGy condition compared to RNDxREGy. While both models successfully account for key aspects of auditory statistical learning, they fall short in explaining the neural dynamics underlying abrupt changes in auditory structure. These findings suggest that the brain may employ specific heuristics or biologically grounded mechanisms to detect change points in background auditory scenes—strategies not yet

incorporated into current computational models. Future work that integrates such constraints into predictive frameworks may yield a more accurate and realistic understanding of auditory perception. Additionally, the pronounced impact of prior context on learning speed underscores the importance of accounting for statistical history—even during passive listening—to more fully understand how the brain functions in dynamic environments.

4. Chapter 4: How Dynamic, Task-Irrelevant Auditory Statistical Changes Shape Visual Memory

4.1 Summary

It is well established that continuous experiences are segmented into smaller "events" and stored in memory as discrete units. This study investigated whether changes in background (behaviourally irrelevant) sound statistics could create such event boundaries and influence long-term memory for concurrently presented visual events. Specifically, I focused on transitions between regular (REG) and random (RND) tone-pip sequences. It was predicted that transitions from REG to RND would form stronger event boundaries than the reverse, as the abrupt loss of structure could serve as a salient signal of environmental change. I hypothesised that such boundaries would lead to: (1) impaired temporal order memory for visual items spanning the boundary compared to items within the same context; (2) longer subjective temporal distance estimates for boundary-spanning item pairs; and (3) enhanced item recognition for visual events experienced concurrently with the boundary. In Experiment 1, I examined predictions (1) and (2). The results supported (1): transitions in background sound statistics—regardless of direction—were sufficient to induce memory segmentation, leading to impaired temporal order memory across the boundary. However, no consistent effects were observed for subjective time estimates. Experiment 2 tested prediction (3), but did not reveal significant enhancement in item memory at boundary moments. Overall, although the effects were weaker than those typically observed with task-relevant boundary signals, these findings demonstrate that changes in task-irrelevant auditory statistics can influence memory organisation in a broader, cross-modal context.

4.2 Introduction

So far, I have focused on how dynamically changing sound statistics are represented in the brain. The two EEG studies discussed above, along with related research, demonstrate that the brain is highly sensitive to shifts in auditory statistics—even when these sounds are presented as task-irrelevant background stimuli. Notably, transitions from highly predictable sequences elicit a sharp drop in sustained neural activity and activate the pupil-linked locus coeruleus–norepinephrine (LC-NE) system (Barascud et al., 2016; Basgol et al., 2025; Bianco et al., 2025; Hu et al., 2024; Magami et al., 2025; Zhao, Chait, et al., 2019; Zhao et al., 2025).

Given that the amplitude drop has been hypothesised to reflect model disengagement—marked by reduced activity in a network involving the auditory cortex, inferior frontal gyrus, and hippocampus (Bianco et al., 2025)—and that norepinephrine (NE) is known to facilitate the processing of bottom-up sensory signals (Gelbard-Sagiv et al., 2018; T. H. Lee et al., 2018; Nassar et al., 2012; Sara, 2009; Sara & Bouret, 2012), these dynamics are considered to reflect a reset of the brain's predictive model, followed by an exploratory phase aimed at constructing a new model adapted to the updated sensory environment.

Despite this evidence of rapid, state-level neural reconfiguration, it remains unclear how such changes impact other ongoing cognitive processes. In everyday life, we often encounter auditory environments passively, while our attention is directed toward a different, task-relevant goal. Yet, we know surprisingly little about how—or whether—these shifts in neural state, triggered by background statistical changes, influence performance on the tasks we are actively engaged in.

One potential consequence of changes in background sound is their influence on how concurrent experiences are segmented into discrete episodes. Interestingly, the neural responses triggered by changes in auditory statistics closely resemble those observed during episode formation in memory.

A central question in memory research is how continuous streams of experience are segmented into distinct episodes—how the brain determines which elements belong together within a single episode and which are treated as separate. One influential account, Event Segmentation Theory (EST), proposes that we maintain mental models to represent ongoing experience, and that these models are updated at event boundaries—points where meaningful changes in the environment are detected (Reynolds et al., 2007; Zacks et al., 2001, 2007). Such event boundaries have been shown to emerge from various stimulus features, including shifts in time and space (Ezzyat & Davachi, 2011; Horner et al., 2016), emotion (Clewett & McClay, 2025; McClay et al., 2023) and perceptual features of the stimuli (DuBrow & Davachi, 2013, 2014; Heusser et al., 2018; Pu et al., 2022).

A growing body of research inspired by EST has shown that event boundaries are not only moments of perceptual reorganisation but also play a critical role in how information is encoded and later retrieved from memory; Items within the same episode tend to be bound together, whereas items across episodes are more likely to be kept separate (Clewett et al., 2020, 2025; Clewett & McClay, 2025; DuBrow & Davachi, 2013, 2014; Heusser et al., 2018; Horner et al., 2016; McClay et al., 2023; Pu et al., 2022; Raccah et al., 2023; Rouhani et al., 2020; Sols et al., 2017). Recent work by Clewett and colleagues (2020, 2025) highlights the key role of the LC-NE system in memory formation at event boundaries. Clewett et al. (2020) demonstrated that event boundaries were accompanied by transient increases in pupil-linked LC-NE activity, and the magnitude of this response predicted the degree of memory separation across the boundary in the later memory test. Extending these findings, Clewett et al. (2025) used fMRI to show that boundary-evoked activity in the LC was associated with increased temporal pattern separation of items across the boundary in the left dentate gyrus (DG) of the hippocampus. Moreover, the strength of LC activation predicted the degree of memory separation across the boundary in the later memory test. These findings suggest a mechanistic role for LC-NE activity in reconfiguring hippocampal networks at event boundaries,

supporting a memory "reset" process that promotes the formation of discrete episodic segments.

This memory reset process following the detection of an event boundary strikingly parallels the proposed prediction model reset that occurs in response to changes in background sound statistics. This raises the intriguing possibility that shifts in background auditory statistics may contribute to the formation of episodes in a broader, multimodal context—such that changes in these statistics act as boundaries that segment concurrently presented experiences. Prior studies have shown that during passive listening, rapid shifts in brain state—marked by abrupt drops in sustained neural response and activation of the pupil-linked LC-NE system—do not occur in response to all statistical transitions. Rather, such neural signatures emerge only when the transition occurs from a highly predictable sequence (Barascud et al., 2016; Basgol et al., 2025; Bianco et al., 2025; Magami et al., 2025; Zhao, Chait, et al., 2019; Zhao et al., 2025). This suggests that background sound transitions may trigger event boundaries only when the change involves a violation of precise prediction.

While EST has primarily been studied in the visual domain, there is growing evidence that its principles extend to auditory and multimodal contexts (Clewett et al., 2020, 2025; Clewett & McClay, 2025; McClay et al., 2023; Raccah et al., 2023). For example, Raccah et al. (2023) exposed participants to sequences of words spoken by either male or female voices, with the task of encoding the order of the items. They indicated that a change in speaker (from male to female or vice versa) functioned as an event boundary, segmenting the ongoing stream and influencing memory for the sequence.

However, most auditory studies have focused on task-relevant sounds, leaving open the question of whether task-irrelevant background sounds can similarly drive event segmentation processes (but see McClay et al., 2023). In this chapter, I explore whether changes in the statistical structure of task-irrelevant background sounds can influence performance on an ongoing task, using the framework of event segmentation theory.

4.3 Experiment 1

In this experiment, I focus on two aspects of memory known to be influenced by event boundaries: **temporal order** and **perceived temporal distance**. It is well established that items experienced within the same context are more likely to be bound together in memory, whereas items that span an event boundary are more likely to be stored as part of separate episodes, which is also reflected in greater representational dissimilarity of those items in the hippocampus (Clewett et al., 2019; DuBrow & Davachi, 2014; Ezzyat & Davachi, 2014). This segmentation disrupts the ability to retrieve the precise order of items saved in different events and also leads to an expansion in the perceived temporal distance between them (Clewett et al., 2020, 2025; Clewett & McClay, 2025; DuBrow & Davachi, 2013, 2014; Heusser et al., 2018; Horner et al., 2016; McClay et al., 2023; Pu et al., 2022; Raccah et al., 2023; Rouhani et al., 2020; Sols et al., 2017).

Here, I investigated whether changes in the statistical structure of background sounds can induce an event boundary effect on memory for visually presented items. Specifically, I manipulated transitions between regular (REG) and random (RND) sound sequences and hypothesised that only the REG-to-RND (REGRND) transition would trigger a boundary effect. This prediction is based on previous findings showing that REGRND—but not RND-to-REG—transitions elicit pupil-linked LC-NE system activation (Basgol et al., 2025; Zhao, Chait, et al., 2019), which is considered to support boundary-related memory segmentation (Clewett et al., 2020, 2025).

In parallel with the main research question, this study also explored whether differences in the predictability of background sounds induces different level of arousal. Prior research on sustained neural responses has shown that predictable (REG) and less predictable (RND) sound sequences elicit different levels of sustained activity—often interpreted as reflecting the brain's confidence in its sensory predictions, or precision (Barascud et al., 2016;

Southwell et al., 2017; Magami et al., 2025; see also Chapter 3). Highly unpredictable environments are generally perceived as more stressful or potentially threatening (Amat et al., 2005; de Berker et al., 2016; Koolhaas et al., 2011; A. Peters et al., 2017), prompting the brain to allocate additional computational resources to resolve the uncertainty (A. Peters et al., 2017). This suggests that RND sequences, compared to REG, may demand more processing resources. Supporting this, a previous pupillometry study reported that RND sequences evoke greater arousal than REG sequences, suggesting that processing RND sequences requires greater cognitive effort (Milne, Zhao, et al., 2021). However, it is important to note that in that study, participants engaged in sound-related tasks, making the auditory sequences task-relevant.

Here, I extend this line of research by examining whether auditory sequences with differing levels of predictability influence arousal even when they are task-irrelevant. To assess this, I measured participants' skin conductance activities. Skin conductance (SC) is a widely used psychophysiological measure that reflects changes in the electrical conductance of the skin due to sweat gland activity (Boucsein, 2012; Dawson et al., 2016; Tronstad et al., 2022). Because sweat glands are innervated exclusively by the sympathetic branch of the autonomic nervous system, SC serves as a sensitive index of sympathetic arousal (Bach, 2014; Boucsein, 2012; Dawson et al., 2016; Tronstad et al., 2022). SC has been extensively used to assess emotional arousal, cognitive effort, and stress. Increases in SC are consistently observed in response to stress-inducing stimuli (Bach et al., 2011; de Berker et al., 2016; Raio et al., 2017), emotionally salient events (Bradley et al., 2008; Ojala & Bach, 2020; Salimpoor et al., 2011; Vinberg et al., 2022), and cognitively demanding tasks (Critchley, 2002; Dawson et al., 2016; Zhang et al., 2012).

There are several notable dissociations between SC and pupil dilation response (PDR), two commonly used physiological measures of arousal. Unlike SC, the pupil size is controlled by a combination of sympathetic and

parasympathetic activity, making it more difficult to isolate the specific contribution of each autonomic branch (Joshi et al., 2016; Joshi & Gold, 2020; Mathôt, 2018; Sirois & Brisson, 2014). Furthermore, SC reflects sympathetic activities mediated by acetylcholine (Ach; Bach, 2014; Boucsein, 2012; Dawson et al., 2016; Tronstad et al., 2022). Notably, this neurotransmitter works as part of parasympathetic activity in the pupil control circuit, with noradrenaline mediating the sympathetic activity (Joshi & Gold, 2020; Sirois & Brisson, 2014). Given these differences, examining SC responses to REG versus RAN stimuli may provide a distinct window into arousal mechanisms that are not captured by PDR.

To summarise, this experiment investigated two questions: (1) whether transitions in background sound statistics can serve as event boundaries that influence memory, and (2) whether differences in the predictability of the background sound influences on the general arousal level.

4.3.1 Methods

4.3.1.1 Auditory stimuli

The stimuli (**Figure *4.1A***) were 200 s long sequences composed of 50 ms tone pips (5 ms raised cosine ramps). Tone frequencies were drawn from a pool of 20 logarithmically spaced values between 222 and 2000 Hz. Each sequence consisted of alternation between regular (REG) and random (RND) patterns. REG patterns were generated by permuting the 20 frequencies from the pool and repeating this sequence to create a regularly repeating pattern. RND patterns were generated by randomly sampling frequencies from the pool with replacement. These REG and RND patterns were played alternately to form a long sequence with 4 ‘transitions’ (e.g., REG-RND-REG-RND-REG). Each of the REG and RND patterns in the sound sequence was unique. The duration of each REG/RND pattern was randomly chosen between 30 s and 50 s (5 s steps) without replacement (the total stimulus duration was always 200 s).

Half of the stimuli began with REG and the other half with RND. A unique sound stimulus was generated for each trial and participant.

To reduce the predictability of the pattern order, 7 out of 15 stimuli (see Procedure) included a switch in the pattern type. For instance, ‘REG-RND-REG-**REG**-REG' was played instead of ‘REG-RND-REG-**RND**-REG'. The switch location was randomly selected. If the pattern switched to REG, a new REG sequence was generated. If it switched to RND, frequencies were drawn from a limited pool, as transitions between RND are undetectable unless the frequency range changes. In this condition, frequencies were selected from a limited pool of 10 frequencies within the middle range of the original pool (397 – 1122 Hz), referred to here as RNDm. Any responses (both behavioural and skin conductance) collected during these switched patterns were not used for subsequent analyses.

4.3.1.2 Visual Stimuli

The visual stimuli consisted of 600 coloured images of everyday objects presented on a white background. Images were selected from Bank of Standardized Stimuli (BOSS; Brodeur et al., 2010) and DinoLab object database (https://mariamh.shinyapps.io/dinolabobjects/) and resized to be 300x300 pixels. Each image was a unique exemplar with a distinct name. Images that were highly arousing, such as food, military equipment, and animals were not included in this image set.

4.3.1.3 Audio-visual stimulus presentation timing

Visual stimuli were presented concurrently with auditory stimuli during the encoding session (see Procedure). The onset of the sound was synchronised with the presentation of the first image. The transition of the sound pattern (e.g., REG-RND) was also synchronised with the image onset (**Figure *4.1B***). For REG to RND transition, the change is immediately perceptible at the onset of the RND segment; thus, a new image was presented

at this point. In contrast, for RND-REG transition, the change only becomes discernible during the second cycle of the REG pattern, as the first cycle is acoustically identical to a RND sequence. Therefore, a new image was presented one cycle after (1 second after REG onset).

4.3.1.4 Procedure

Participants were seated in an acoustically shielded room (IAC triple-walled sound attenuating booth). The experiment comprised four phases: a resting period, an encoding session, a distraction task, and a memory test (**Figure *4.1C***). This 'block' was repeated 15 times, with short breaks in between blocks. The first block served as a practice session; data from this block were excluded from subsequent analyses.

The resting period lasted 1–2 minutes and continued until the participant's skin conductance signal stabilised. To facilitate this, participants were instructed to relax and breathe slowly. Once it had stabilised, the skin conductance signal was recorded as baseline activity.

During the encoding session, 40 unique images were presented on the screen while the auditory stimulus played concurrently. Each image appeared for 2.5 s, with a 2.5 s inter-stimulus interval (ISI) during which a fixation cross was displayed. The order of image and sound presentation was randomised across participants. Participants were instructed to memorise the sequence of images and were specifically encouraged to do so by creating narratives that linked successive items. They were asked to press the 'Enter' key with their dominant (right) hand once they had finished encoding each image. The timing of the button press did not affect the duration of image presentation or the ISI.

After completing the encoding session, participants completed a 45-second memory disruption task. In this task, 45 arrow images were presented in rapid succession. Participants were instructed to identify the direction of each arrow as quickly and accurately as possible. Each arrow was displayed for 500

ms, followed by a 500 ms ISI. Feedback on both hits and misses was provided at the end of the task.

Finally, the memory test session assessed two aspects of episodic memory: temporal order and temporal distance. Participants were shown 15 pairs of objects from the encoding session, presented one pair at a time. Each pair consisted of items that had either been presented within the same sound pattern (e.g., within the same REG) or across two adjacent patterns (e.g., REG and RND; **Figure *4.1D***). In each trial, participants first indicated which object appeared more recently during the encoding session. Following this response, the same pair remained on the screen, and participants rated the perceived temporal distance between the two items. Ratings were made on a four-point scale: very close, close, far, or very far apart. Importantly, all item pairs were separated by three intervening items during encoding, resulting in a constant objective distance. Thus, any variation in perceived temporal distance reflects purely subjective memory judgments. Each response was constrained to an 8-second time window, and the display advanced immediately after a button press. No auditory stimuli were presented during this session. At the end of the test, participants received feedback on their temporal order accuracy. The order of item pair presentations and their left/right screen positions were fully randomised.

Visual stimuli were presented on a computer screen positioned approximately 90 cm from the participant. Auditory stimuli were presented diotically through headphones (HD558, Sennheiser) via a Fireface UC sound card (RME) at a comfortable listening level (adjusted by each participant). Stimulus presentations were controlled with the Psychtoolbox package (Psychophysics Toolbox Version 3) in MATLAB (2019b The MathWorks, Inc.).

4.3.1.5 Recording and data processing of skin conductance data

Skin conductance (SC) was recorded using the Biosemi system (Biosemi ActiveTwo AD-box ADC-17, Biosemi, Netherlands) at a sampling rate of 2048 Hz. A constant current of 1 µA was applied via two flat Ag/AgCl electrodes

(Biosemi), filled with 0.5% NaCl gel (GEL 101, Biopac; Hygge & Hugdahl, 1985). These electrodes were attached to the participant's left (non-dominant) index and middle fingertips using double-sided adhesive pads and tape. Flat-type CMS and DRL electrodes (Biosemi) were placed approximately two inches apart on the back of the same hand.

Data were down-sampled to 32 Hz and processed using the Fieldtrip (http://www.fieldtriptoolbox.org/) toolbox for MATLAB (2018a, MathWorks) and Ledalab (v 3.4.9, http://www.ledalab.de/). Movement artefacts were visually identified and corrected using spline interpolation. Phasic activities to the stimuli were extracted via Continuous Decomposition Analysis (CDA; Benedek & Kaernbach, 2010) to find non-responding blocks. Blocks with zero event-evoked phasic activities were excluded as outliers, resulting in an average removal of 0.04% of blocks per participant. Also, only the first half of experimental blocks (Blocks 2–8; approximately the first hour of recording) were included in the SC analysis to minimise effects of signal drift over time.

To examine the relationship between SC activity and temporal memory performance, 200-second SC time series from each block were baseline-corrected relative to the resting period from the corresponding block. The final 10 seconds of the resting period were averaged and used as the baseline. Data were averaged across all blocks per participant to obtain an overall measure of physiological arousal during the task. These data were divided into high and low performer groups based on a median split of their average temporal order memory scores. These scores were computed using data from Blocks 2–8, including only the REG and RND (non-transition) conditions, which provided a baseline measure of task performance unaffected by statistical transitions.

To assess the effect of background sound predictability (REG vs RND) on SC activity, z-transformed SC time series (200 s per block) were segmented into 30-second epochs time-locked to the sound onset. Only the first auditory pattern in each block was analysed to eliminate potential carryover effects from preceding sound contexts, ensuring that SC responses reflected only the

response to sound condition of interest. Epochs were baseline-corrected relative to 0–1 s post-sound onset, a choice given the slow rise time of SC responses, which typically emerge around 1 s following stimulus onset (Dawson 2016; Boucsein 2012).

4.3.1.6 Data processing of behavioural measures

For the behavioural analyses, several measures were collected: response time during the encoding session; response time and accuracy from the temporal order memory test; and response time and distance judgments from the temporal distance memory test. For all analyses, responses to the first image of each block were excluded, as this item was presented at the transition from silence—a point that may act as a memory boundary. In the temporal distance memory test, distance ratings were converted to a numerical scale ranging from 1 (very close) to 4 (very far). The scores were then z-transformed per participant to standardise the data.

4.3.1.7 Statistical analysis

For the behavioural data, difference between conditions were first evaluated using repeated-measure ANOVA. Pairwise t-tests were conducted for post-hoc comparisons when the ANOVA revealed a significant effect ($p<.05$). These t-tests were planned, so no adjustments were made for multiple comparisons.

For the statistical evaluation for the time series SC data, the difference between sound conditions were calculated for each participant. This difference was then subjected to bootstrap resampling (Efron & Tibshirani, 1994). The difference between conditions was considered significant if the proportion of bootstrap iterations falling above or below zero exceeded 95% ($p<.05$).

4.3.1.8 Participants

Forty-three paid, right-handed participants were recruited for Experiment 1. One participant with dyslexia was excluded from all analyses. For the behavioural analysis, item order memory performance was assessed, and

blocks in which participants made errors on more than half of the trials were flagged. The number of such blocks was tallied per participant, and those having flagged blocks more than two standard deviations from the mean (four participants) were excluded. This resulted in a final behavioural sample of 38 participants (29 females; mean age = 23.0, ± 3.8). For skin conductance analysis, three participants were excluded as non-responders due to a lack of stimulus-evoked responses, and one additional participant was excluded due to a technical recording failure. The remaining dataset for this analysis also included 38 participants (29 females; mean age = 22.6, ± 3.6). All participants reported no history of hearing or neurological disorders. All experimental procedures were approved by the research ethics committee of University College London, and written informed consent was obtained from each participant.

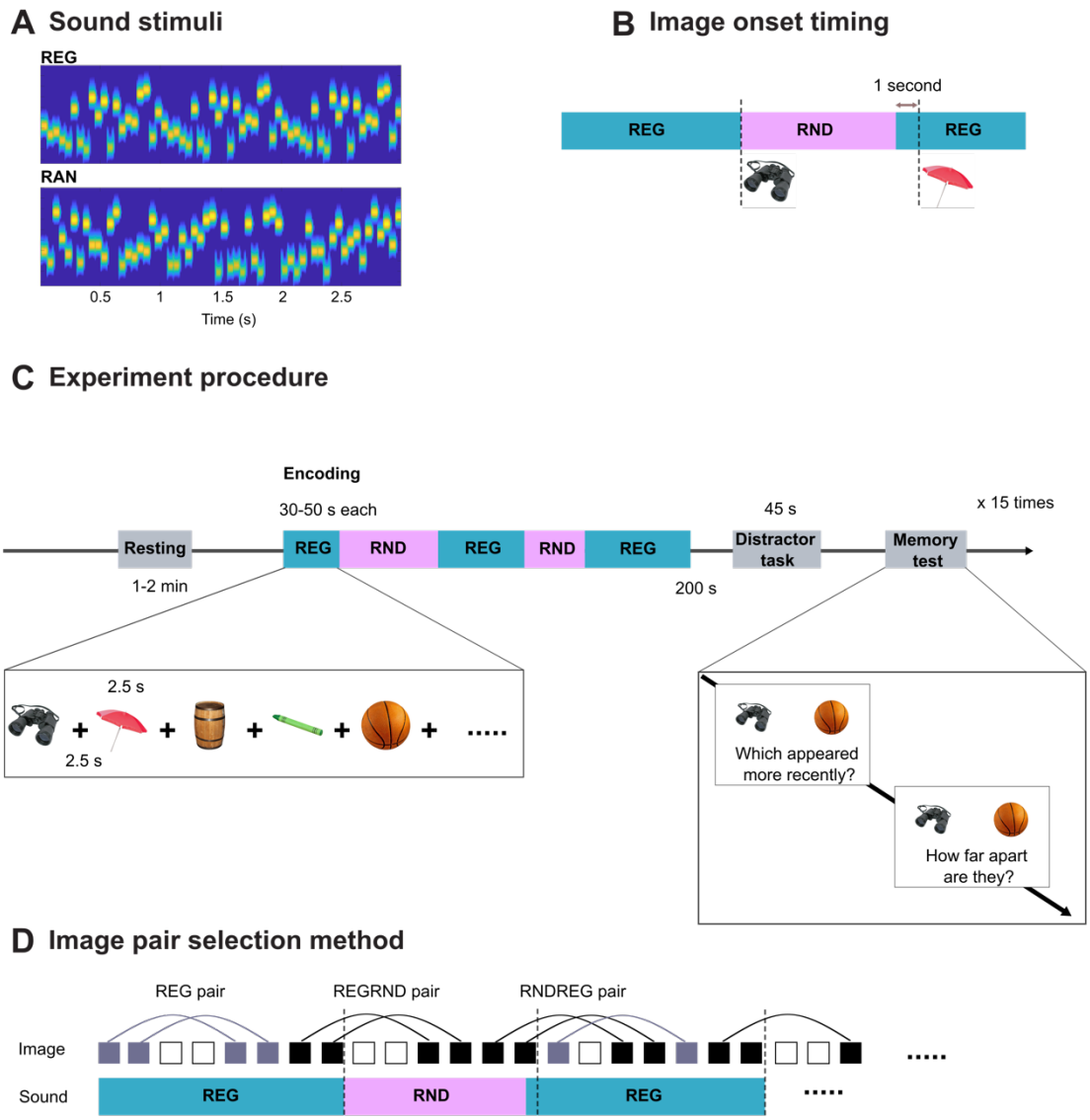


Figure 4.1 Stimuli and task schematics.

[A] Spectrograms depicting example stimuli for REG and RND, showing only the initial 3 s. **[B]** Image onset timings relative to the sound stimuli. For RNDREG transition, the image is presented after one cycle (i.e. 1 s) of REG. **[C]** Experiment procedure. The experiment contained a resting period, an encoding phase, a distractor task, and a memory test. Participants repeated this process 15 times. **[D]** Image pairs used in the memory test are selected from the encoding phase.

All item pairs were separated by three intervening items. The object images in this figure are credited to: https://mariamh.shinyapps.io/dinolabobjects/ and Brodeur et al. (2010).

4.3.2 Results

4.3.2.1 Encoding RT

I first examined how background sound conditions influenced the speed of image encoding. A significant main effect of sound condition on response time was observed ($F(3, 105) = 2.99$, $p = 0.03$, $\eta^2p = 0.08$), with participants responding significantly faster to images presented during REG sounds compared to those presented during RND sounds ($t(35) = -2.25$, $p = 0.03$, $d = -0.37$), as well as to those presented immediately after a transition from RND to REG ($t(35) = -2.80$, $p = 0.008$, $d = -0.47$; **Figure *4.2A***). This pattern generally aligns with previous findings showing that task-irrelevant regular auditory patterns can enhance performance and speed up responses to attended tasks compared to random patterns (Southwell et al., 2017). In the context of boundary theory, however, studies using similar paradigms to investigate event boundaries rarely report encoding response times (McClay et al., 2023; Pu et al., 2022; Raccah et al., 2023), which limits direct comparisons to the current results.

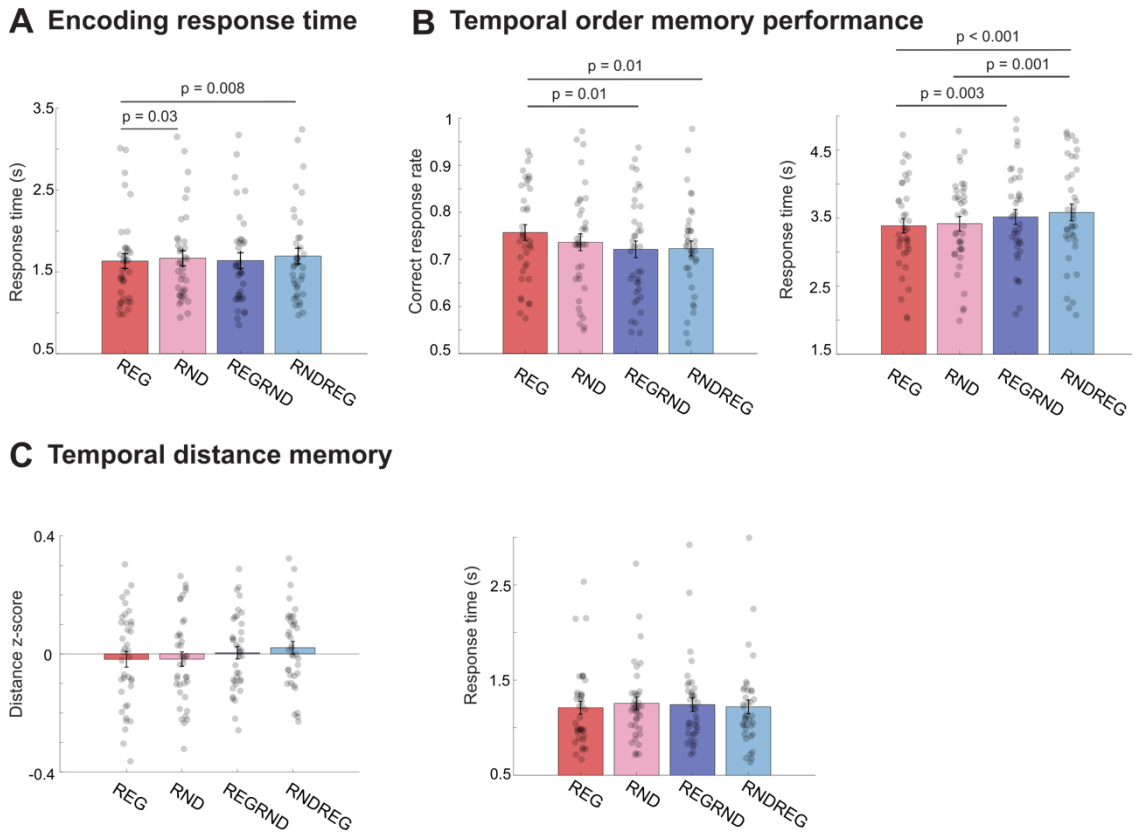


Figure 4.2 Behavioural results of the influence of the background sound on the ongoing visual task.

[A] Average response time for the item encoding. **[B]** Correct response rate (left) and the response time of the accurate trials (right) of the temporal order memory test. **[C]** z-transformed distance score (left) and the response time (right) of the temporal distance memory test. In all figures, the error bars represent the s.e.m. Overlaid dots represent individual participants. Significant results of the pairwise t-tests were plotted on the top of the bar plots.

4.3.2.2 Effects of transition in sound statistics on temporal memory

To investigate whether and how background sound modulates the temporal structure of memory, two aspects of episodic memory were tested: temporal order and temporal distance.

For temporal order memory, both the accuracy of responses and the response times for correct answers showed significant main effects of sound condition (accuracy: $F(3, 111) = 3.02$, $p = 0.03$, $\eta^2p = 0.08$; response time: $F(3, 111) = 7.67$, $p < 0.001$, $\eta^2p = 0.17$; **Figure *4.2B***). Post-hoc comparisons revealed that temporal order memory was significantly impaired in both transition conditions compared to the REG condition (REG vs REGRND: $t(37) = 2.71$, $p = 0.01$, $d = 0.44$; REG vs RNDREG: $t(37) = 2.73$, $p = 0.01$, $d = 0.44$). Additionally, participants took longer to select the correct response in the transition conditions than in the no-transition conditions (REG vs REGRND: $t(37) = -3.19$, $p = 0.003$, $d = -0.52$; REG vs RNDREG: $t(37) = -4.77$, $p < 0.001$, $d = -0.77$; RND vs RNDREG: $t(37) = -3.50$, $p = 0.001$, $d = -0.57$; **Figure *4.2B***). No significant differences were found between the REGRAN and RANREG conditions in any comparison. These results suggest that changes in background sound statistics—regardless of transition direction—disrupt temporal order memory for items presented across different sound contexts, implying that such transitions may induce the formation of event boundaries. Importantly, this impairment is unlikely to be driven by distraction from the abrupt sound change at the transition itself. As shown in **Figure *4.1D***, item pairs tested in the transition conditions were selected to span the boundary, with one item presented before the transition and the other at least 10 seconds after it.

For temporal distance memory, a subjective time expansion effect was expected for item pairs that spanned an event boundary, despite objective temporal distance between item pairs was identical across all trials. However, contrary to this expectation, temporal distance memory did not show the expansion effect associated with event boundaries (distance score: $F(3, 111) = 0.49$, $p = 0.69$, $\eta^2p = 0.01$; response time: $F(3,111) = 1.63$, $p = 0.19$, $\eta^2p = 0.04$; **Figure *4.2C***).

Overall, the results suggest that transitions in task-irrelevant sound statistics form event boundaries in memory. However, the strength of the boundary may be weak due to various reasons including task-irrelevance. Indeed, the failure to show the distance effect, that is commonly associated with smaller effect sizes than the order effect (Clewett et al., 2020; McClay et al., 2023) would be consistent with this interpretation.

4.3.2.3 Memory performance is reflected in skin conductance activity.

To examine whether individual skin conductance activity throughout the experiment reflects task performance, as measured by the temporal order memory score, we averaged the baseline-corrected skin conductance signals across blocks for each participant and compared these values between high and low performers (see Method). Bootstrap resampling revealed a significant difference between the two groups, with consistently higher skin conductance observed in high performers (**Figure *4.3A***). This finding aligns with the well-established idea that physiological arousal reflects the level of task engagement and performance (Aston-Jones et al., 1999; de Gee et al., 2024; Waschke et al., 2019).

4.3.2.4 Background sound predictability induced different skin conductance activity.

Next, I examined whether background sounds with differing levels of predictability evoked different levels of skin conductance activity. To eliminate potential carryover effects from preceding sound patterns, I only analysed the REG and RND sequences presented at the start of each block (7 sequences per participant). Bootstrap resampling revealed significantly higher skin conductance activity during the presentation of RND compared to REG sequences (**Figure *4.3B***).

Interestingly, when participants were divided into high and low performers based on their temporal order memory scores (see Method), this difference was significant only in the low-performing group (**Figure *4.3B***). Taken together with the previous findings, this suggests that high performers

maintained elevated general arousal due to sustained task engagement, leaving little room for the influence of background sound to modulate arousal levels. In contrast, low performers—who consistently showed lower task-related arousal—may have been more susceptible to the effect of background sounds. This susceptibility allowed the differential arousal evoked by REG and RND patterns to become apparent in their physiological responses.

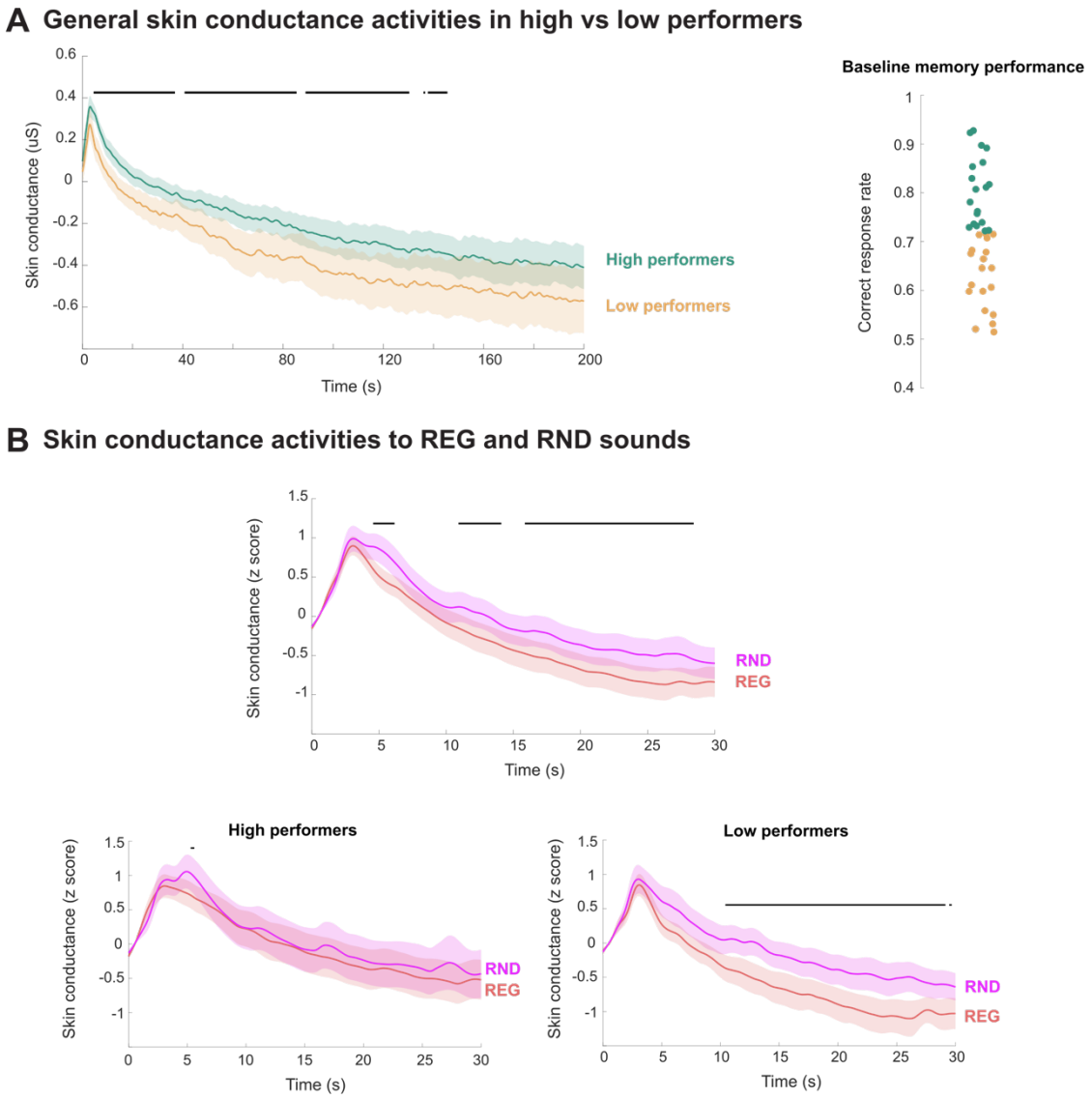


Figure 4.3 Skin conductance activity measured during the encoding session.

[A] Left: Group-averaged skin conductance. Participants are divided into two groups based on their performance on the temporal order memory test (REG and RND conditions only, see Methods). Skin conductance data are baseline corrected relative to the resting period immediately preceding each block. Shaded areas represent twice the SEM. Significant differences ($p<.05$) between two performer groups are indicated by bold horizontal lines above the skin conductance traces. Right: Temporal order memory score for high and low performers. Baseline performance is calculated as the average across non-transition conditions (REG and RND, see Method) and plotted. Each dot represents an individual participant: green dots indicate high performers and orange dots indicate low performers. **[B]** Top: Group-averaged skin conductance evoked during REG and RND sound presentation. Data are baseline corrected to 0-1 s. Shaded areas represent twice the SEM. Significant differences ($p<.05$) between two performer groups are indicated by bold horizontal lines above the skin conductance traces. Bottom: same as top, but participants were separated into high and low performers based on their temporal order memory score.

4.3.3 Discussion

This experiment investigated whether and how task-irrelevant background sounds—specifically, changes in their statistical structure—can influence visual memory encoding and retrieval. Behavioural measures combined with skin conductance recordings revealed that (1) shifts in background sound statistics can induce event boundaries, and (2) the

predictability of background sounds modulates general arousal levels, even during tasks unrelated to the auditory input.

4.3.3.1 Transition in background sound statistics forms an event boundary

In this experiment, I manipulated behaviourally irrelevant background sound statistics to examine whether a transition in auditory context could form a boundary in memory. I specifically hypothesised that the REGRAN transition would induce a boundary, based on prior evidence suggesting that REGRAN—but not RANREG—activates the pupil-linked locus coeruleus–noradrenaline (LC-NE) system (Basgol et al., 2025; Zhao, Chait, et al., 2019). However, our results revealed impaired temporal order memory in both transition conditions when compared to the continuous REG condition, indicating that a boundary was formed regardless of the direction of the statistical change.

This finding aligns with the concept of event boundaries as markers of contextual shifts (Clewett et al., 2019; Reynolds et al., 2007; Zacks et al., 2001, 2007). In this framework, any transition in sound statistics—whether from regular to random or vice versa—signals a meaningful change in context and is therefore sufficient to induce memory segmentation. This suggests that both REG and RND sequences were perceived as stable, distinct contexts, similar to those previously observed in other domains such as semantic categories (e.g., items belonging to the same conceptual group; DuBrow & Davachi, 2013, 2014; Manning & Kahana, 2012), emotional states (e.g., items encoded during a specific mood; Clewett & McClay, 2025; McClay et al., 2023), or spatial locations (e.g., items experienced within the same environment; Horner et al., 2016). Notably, there was no significant difference in memory performance between the REG and RND conditions—that is, when items were experienced within a stable REG or RND context (**Figure *4.2B***). However, the memory impairment observed in the transition conditions was primarily evident in comparison to the REG context (**Figure *4.2B***). This suggests that while both REG and RND sequences can serve as stable contexts, the REG context may

offer greater contextual coherence, likely due to its deterministic and highly predictable structure.

Although previous work did not observe activation of the pupil-linked LC-NE system in response to RND-to-REG transitions (Basgol et al., 2025; Zhao, Chait, et al., 2019), this does not preclude the possibility that such activation occurred in the present study, given substantial differences in experimental design. Most notably, the duration of each auditory pattern in our paradigm was considerably longer (30–50 seconds) compared to the 2.5–3.5 seconds used in the earlier work. This extended exposure may have allowed to form more stable predictions about the ongoing auditory context, even in the case of RND sequences. With sufficient exposure, the brain may develop a statistical model of the auditory environment—such as the expectation of a uniform distribution of tones—which, when violated by a sudden shift in regularity, could still trigger LC-NE engagement as part of a context-resetting mechanism. Alternatively, it remains possible that LC-NE activation is not essential for boundary formation.

Recent work by Clewett et al. (2025) has highlighted a potential role of the LC in resetting hippocampal memory representations following contextual shifts. While their findings are correlational and the causal role of the LC-NE system in boundary-induced memory segmentation remains to be established, their conclusions are consistent with broader literature on change-point detection. These frameworks propose that changes in environment prompt the brain to reset its internal model about the current environment and begin gathering new evidence—processes in which LC-NE activity is thought to play a critical role (Lawson et al., 2021; Nassar et al., 2010, 2012; Skerritt-Davis & Elhilali, 2021a, 2021b). Future research should investigate whether LC-NE activation is a necessary condition for boundary-driven memory effects, as current evidence is based on limited boundary types and task designs (Clewett et al., 2020, 2025).

While no significant difference was observed in temporal distance memory performance (**Figure *4.2C***), it is important to note that this measure is

relatively less established and tends to yield smaller effect sizes (Clewett et al., 2020; Clewett & McClay, 2025; McClay et al., 2023). Indeed, several studies have either omitted reporting these results despite collecting the data or reported inconsistent findings when using distance-based memory metrics (Clewett et al., 2025; Clewett & McClay, 2025; Rouhani et al., 2020).

Nevertheless, the findings demonstrate that, even in the absence of task relevance, background sound statistics can reliably establish contextual frameworks that influence the encoding of concurrently presented visual information.

4.3.3.2 Background sound predictability influences general arousal

Skin conductance revealed higher arousal levels in the RND condition compared to the REG condition (**Figure *4.3B***). This aligns with findings from Milne, Zhao, et al. (2021), who reported that the tonic component of the pupil dilation response (PDR) was higher for the RND sound than the REG sound. They interpreted this as reflecting the sustained computational demand required to process unpredictable input, in contrast to REG sequences, where PDR decreased once the pattern was learned—indicative of reduced processing effort.

Consistent with this interpretation, behavioural results from the present experiment showed slower response times in the encoding task under the RND condition (**Figure *4.2A***), suggesting that unpredictable background sounds could not be easily suppressed and interfered more with task performance. This is in line with prior work by Southwell et al. (2017), who demonstrated the distracting effect of RND versus REG sequences. However, it is also possible that the observed effect reflects facilitation by REG sequences rather than interference from RND. Without a silent control condition, it remains difficult to determine whether REG improves performance, RND disrupts it, or both. Future work incorporating a silence baseline will be necessary to disentangle these possibilities. Take together, these findings support the broader notion of expectation suppression, whereby predictable stimuli elicit reduced neural and

cognitive responses, reflecting increased processing efficiency (de Lange et al., 2018).

Although differences in skin conductance and response times were observed during the encoding stage, these effects did not translate into measurable differences in subsequent memory performance (**Figure *4.2B***). This finding aligns with the finding from Clewett et al. (2025), who reported that the relationship between LC activity during encoding and the later memory performance was specific to boundary trials and absent during stable contexts. This suggests that LC-driven memory modulation occurs selectively during moments of contextual change, when internal representations are updated and encoded as novel events. While skin conductance is not a direct index of LC-NE activity, the absence of a memory effect, despite observed differences in SC between REG and RND conditions during encoding, suggests that the sound-induced arousal differences in this experiment were too subtle to influence memory retrieval performance.

4.3.3.3 How to interpret the skin conductance activity?

The SC activity observed in this experiment resembled previously reported PDR profiles (Milne, Zhao, et al., 2021), prompting the question of whether these effects reflect a shared arousal mechanism. However, the interpretation is more nuanced. Although both SC and PDR are associated with arousal, they are mediated by different neurophysiological pathways: SC is predominantly driven by acetylcholine (ACh)-mediated activation of sympathetic sweat glands, whereas PDR reflects the mixed responses from sympathetic and parasympathetic systems (Bach, 2014; Boucsein, 2012; Dawson et al., 2016; Joshi & Gold, 2020; Sirois & Brisson, 2014; Tronstad et al., 2022). Notably, the comparison of SC and PDR activities induced by fear conditioning and emotional stimuli have mixed results; some find some level of correlation but none of the studies are strong enough to state that they are derived from a common neural source (Bradley et al., 2008; Korn et al., 2017; Leuchs et al., 2019). Although further investigation is needed to clarify the relationship

between SC and PDR as markers of arousal, the similarity in their response patterns suggests that background sound predictability may engage both cholinergic and noradrenergic branches of the sympathetic nervous system. This pattern may reflect a coordinated, multi-system arousal response involving both central and peripheral components of the autonomic nervous system.

In this study, the skin conductance activity was reported across conditions without dissociating skin conductance level (SCL) and skin conductance responses (SCR). While SC is often analysed by separating tonic (SCL) and phasic (SCR) components, this distinction was not feasible in the present paradigm due to the continuous nature of the multimodal stimulus presentation. The absence of discrete, well-defined stimulus onsets made it conceptually challenging to isolate event-evoked responses from ongoing tonic activity. Notably, our analysis (**Figure *4.3***) revealed that the observed SC differences were primarily driven by low-frequency components, with minimal contribution from high-frequency phasic activity. While the full signal was retained in the analysis and visualisation for transparency, it is likely that the condition differences primarily reflect variations in the tonic component.

Overall, this study provides evidence that changes in the task-irrelevant background sound statistics can influence boundary formation. While McClay et al. (2023) previously demonstrated that background sound can structure event boundary, their stimuli were intentionally composed to evoke distinct emotional states (e.g., sadness, calmness), meaning that boundaries were defined by complex combinations of acoustic features associated with emotion. In contrast, the present findings extend this line of work by showing that even in the absence of emotional shifts, mere changes in statistical regularity are sufficient to create contextual boundaries. This suggests a broader role for background auditory structure in memory organisation, beyond the influence of emotional arousal. Remarkably, in the current paradigm, all acoustic features were held constant except for tone order—yet this alone was sufficient to modulate contextual segmentation. Furthermore, building on Milne, Zhao, et al. (2021),

skin conductance measures indicated that the predictability of the background sound sequence modulates arousal levels even when the auditory input is entirely task-irrelevant. Together, these findings highlight the influence of dynamic, task-irrelevant auditory statistics on memory and arousal, underscoring the importance of studying 'background' computations to fully understand how we process and remember information in naturalistic environments where multiple streams of information coexist.

4.4 Experiment 2

In this experiment, I investigated another aspect of the boundary effect on memory: recognition performance for incidentally experienced items. Event Segmentation Theory suggests that boundary information is processed more deeply in order to construct a new event model (Swallow et al., 2009; Zacks et al., 2007; Zacks & Swallow, 2007). Consistent with this, numerous studies have reported enhanced recognition for items presented at event boundaries compared to those encountered within continuous episodes (McClay et al., 2023; Pettijohn et al., 2016; Radvansky et al., 2014; Rouhani et al., 2020; Swallow et al., 2009).

These findings raise the possibility that changes in background sound statistics—shown in Experiment 1 to evoke boundary-like responses—may also enhance recognition of concurrently presented items.

4.4.1 Methods

4.4.1.1 Auditory stimuli

The stimuli (**Figure *4.4A***) were ~9 minutes long sequences composed of 50 ms tone pips (5 ms raised cosine ramps). Tone frequencies were drawn from a pool of 20 logarithmically spaced values between 222 and 2000 Hz. Each sequence consisted of alternation between regular (REG) and random (RND) patterns. REG patterns were generated by randomly selecting 10 frequencies

from the pool without replacement, and this sequence was repeated to create a regularly repeating pattern. RND patterns were generated by randomly sampling frequencies from the pool with replacement. These REG and RND patterns were played alternately (e.g., REG-RND-REG-RND...) to form a long sequence (60 REG and 60 RND patterns appeared in one sequence). Each of the REG and RND patterns in the sound sequence was unique. The duration of each REG/RND pattern was jittered between 3.5 and 5 s (70 – 100 tone-pips). To reduce the predictability of the pattern order, 6 additional REGx-REGy patterns were randomly inserted into the sequence. Both REGx and REGy were created by selecting 10 frequencies from the pool. There are some possible overlaps in the frequencies in REGx and REGy, but the patterns were distinct. The resulting sequence durations varied between 7.7 and 11 min.

Participants were exposed to two long sequences during the experiment: one starting with RND and the other with REG. The presentation order was counterbalanced across participants. A unique sound sequence was generated for each participant.

4.4.1.2 Visual stimuli

The visual stimuli consisted of coloured images used in Clewett et al. (2022), including 125 images of animals and 125 images of tools, presented on a white background. In total, 240 images (120 animals, 120 tools) were used for the main experiment, while 10 images were used for the practice session. Images were resized to be 300x300 pixels. Each image was a unique exemplar with a distinct name.

4.4.1.3 Audio-visual stimulus presentation timing

Visual stimuli were presented concurrently with auditory stimuli during the encoding session (see Procedure). Each image appeared under one of four conditions, based on the timing of its onset within the sound sequence: (1) during the middle of a REG pattern (*midREG*), (2) during the middle of a RND pattern (*midRND*), (3) shortly after a transition from REG to RND (*REGRND*), or (4) shortly after a transition from RND to REG (*RNDREG*).

In the *midREG* and *midRND* conditions, images were presented at the midpoint of the corresponding auditory pattern. For the *REGRND* and *RNDREG* conditions, image onset was aligned with prior neurophysiological findings on transition detection. Barascud et al. (2016) reported that neural responses diverge from baseline approximately 150 ms (3 tones) after a REG-to-RND transition and around 750 ms (15 tones) after a RND-to-REG transition. These latencies were taken as the earliest points at which a transition could be detected and were used to time visual stimulus onset in the respective conditions (150 ms post-transition for *REGRND*, 750 ms post-transition for *RNDREG*).

To ensure even distribution of visual events across the sound sequence, the entire auditory stream was divided into 60 segments, each composed of either a REG–RND or RND–REG sound pair. One image was presented per segment. No images were presented during REGx–REGy segments. Visual stimulus onset and image category were pseudo-randomised according to the following constraints: (1) Each onset condition (midREG, midRND, REGRND, RNDREG) included 15 images. (2) The inter-image-interval was always longer than 3 seconds. (3) Animal and tool images were equally represented across conditions, with each condition containing 40–60% of each category.

4.4.1.4 Procedure

Participants were seated in an acoustically shielded room (IAC triple-walled sound attenuating booth). The experiment comprised four phases: a practice session, two encoding sessions, a distraction task, and a memory test (**Figure *4.4B***). Participants were not informed about the memory test beforehand.

Before the main task, participants completed a brief training session. They viewed 10 images (5 animals and 5 tools) and categorised each by pressing either the "4" or "6" key (counterbalanced). The training continued until participants achieved at least 8 correct responses. No auditory stimuli were presented during training.

Participants then completed two encoding sessions, separated by a one-minute break. Each session lasted approximately 9 minutes and involved the concurrent presentation of visual and auditory stimuli. In each session, participants viewed 60 unique images (30 animals and 30 tools). Images were presented for 500 ms each in a pseudo-randomised order, with no more than three consecutive images from the same category. Participants were instructed to categorise each image (animal vs tool) as quickly and accurately as possible while ignoring the background sound. The key mappings were the same as those used in the training session. Responses made more than 3 seconds after image offset were recorded as null responses. At the end of each session, participants received feedback on their total number of hits.

After completing the encoding sessions, participants completed a 45-second memory disruption task, identical to the one used in Experiment 1. The memory test session commenced directly after the disruption task. To assess participants' expectations about the memory test, a brief questionnaire was administered first (adapted from Dunsmoor et al., 2015). Participants rated their level of surprise using a 5-point Likert scale, ranging from 1 ("Did not expected a memory test at all") to 5 ("Fully expected a memory test").

The recognition task involved the presentation of 240 images: 120 previously viewed during encoding and 120 novel lure images. All images were shown in randomised order. For each image, participants indicated whether it was previously seen ("old") or not ("new"), using a four-point confidence scale: definitely old, maybe old, maybe new, and definitely new. The task was self-paced and conducted without background sound.

All visual stimuli were displayed on a monitor positioned approximately 90 cm from the participant. Auditory stimuli were delivered diotically through Sennheiser HD558 headphones using a Fireface UC (RME) sound card, set to a comfortable listening level (adjusted by each participant). Stimulus presentation was controlled using the Psychtoolbox package (Psychophysics Toolbox Version 3) in MATLAB (2019b, The MathWorks, Inc.).

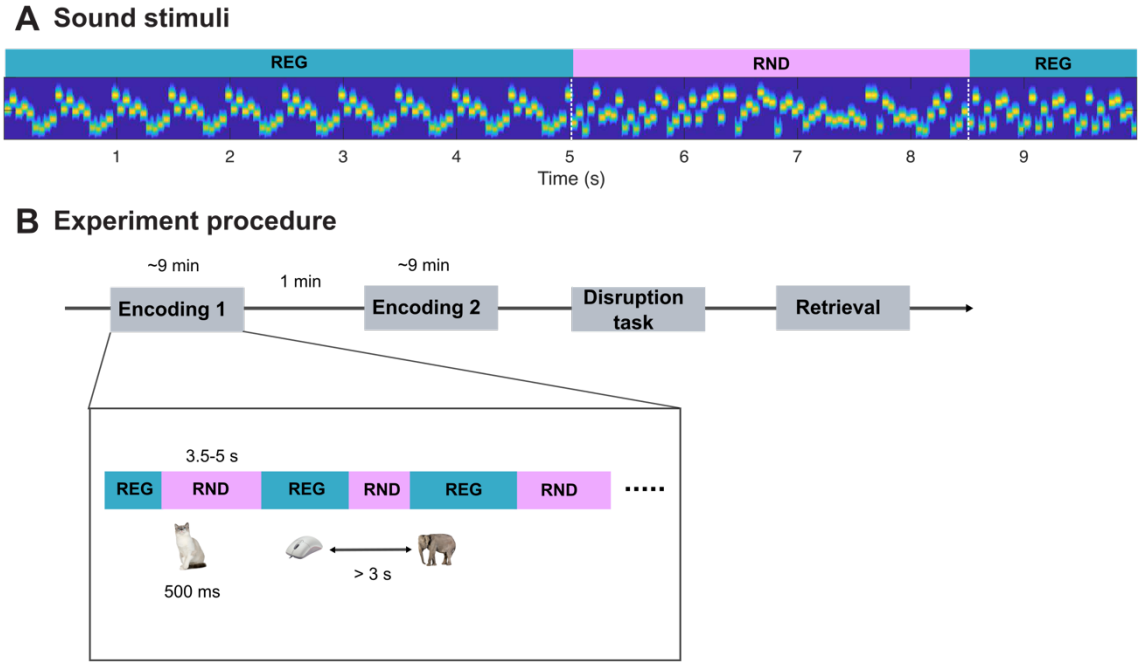


Figure 4.4 Stimuli and task schematics.

[A] Spectrograms depicting example stimuli, showing only the initial 10 s. White dotted lines indicate the timing of transition. Here the sequence transitions from REG, RND, and REG. **[B]** Experiment procedure. The experiment contained two encoding sessions, a memory disruption task, and a memory retrieval session. The object images in this figure are extracted from Clewett et al. (2022).

4.4.1.5 Statistical analysis

To assess the effect of background sound on visual processing, differences between sound conditions were evaluated using a repeated-measures ANOVA. Encoding performance was assessed using both the correct response rate and response time. Memory retrieval performance was quantified using hit rates, calculated both across all trials and restricted to trials with high confidence ratings. The high-confidence hit rate was computed by dividing the

number of correct high-confidence responses by the total number of high-confidence trials. Additionally, participants were split into two groups based on a median split of their overall memory performance scores, as indexed by d', and their responses were analysed separately. In this report, I focus on the standard hit rate rather than d' (except for the performer grouping described above), as the number of false positive was constant across sound conditions. This is because false positives originated from 'new' trials, which were not associated with any background sound condition.

4.4.1.6 Participants

Thirty-four paid participants were recruited for Experiment 2. Three participants were excluded from the analysis: one anticipated the memory test, and two had poor memory performance—either with more than 50% missed items or more than 50% false positives. This resulted in a final behavioural sample of thirty-one participants (24 females; mean age = 22.8, ± 4.7). All participants reported no history of hearing or neurological disorders. All experimental procedures were approved by the research ethics committee of University College London, and written informed consent was obtained from each participant.

4.4.2 Results

4.4.2.1 Encoding performance

I first examined how background sound conditions influenced performance on the animal vs. tool judgment task during the encoding phase. As shown in **Figure *4.5A***, neither accuracy nor response time showed a significant main effect of sound condition (accuracy: $F(3, 90) = 0.94$, $p = 0.42$, $\eta^2_p = 0.03$; response time: $F(3, 90) = 2.24$, $p = 0.09$, $\eta^2_p = 0.07$). Overall, performance was near ceiling, making it difficult to detect any condition-related differences. As an exploratory analysis, the number of missed trials per condition was examined. Although missed trials were rare due to the ceiling effect, I computed the proportion of missed trials per condition, relative to each

participant's total number of misses (e.g., if 8 out of 10 total misses were from the REG condition, the REG proportion would be 0.8). While this analysis is descriptive, the RNDREG condition showed a slightly higher proportion of misses compared to other conditions (**Figure *4.5A***), suggesting the distracting feature of this condition.

4.4.2.2 Retrieval performance

To examine whether background sound influenced the visual encoding of concurrently presented items, memory retrieval performance was compared across sound conditions. As a first step, I confirmed that participants successfully learned the visual items presented during the encoding phase. Collapsing across sound conditions, the hit rate (correctly identifying old images as "old") and false positive rate (incorrectly identifying new images as "old") were calculated (**Figure *4.5B***). The average hit rate roughly aligned with findings from previous studies (Horner et al., 2016; McClay et al., 2023), confirming that the paradigm effectively supported visual memory encoding.

However, comparisons of memory retrieval performance across sound conditions revealed no significant main effect of sound condition on hit rates (**Figure *4.5C***). This was true both when considering all trials and when restricting the analysis to high-confidence responses (all trials: $F(3, 90) = 0.21$, $p = 0.89$, $\eta^2_p = 0.007$; high-confidence trials: $F(3, 90) = 0.14$, $p = 0.94$, $\eta^2_p = 0.005$). Even when participants were divided into high and low performers based on their overall memory performance (see Methods), no significant main effect of the sound condition was observed (high performers: $F(3, 42) = 0.92$, $p = 0.44$, $\eta^2_p = 0.06$; low performers: $F(3, 45) = 0.25$, $p = 0.86$, $\eta^2_p = 0.02$; using all trials). These results suggest that, unlike in Experiment 1, no sound-induced boundary effects on memory performance were observed.

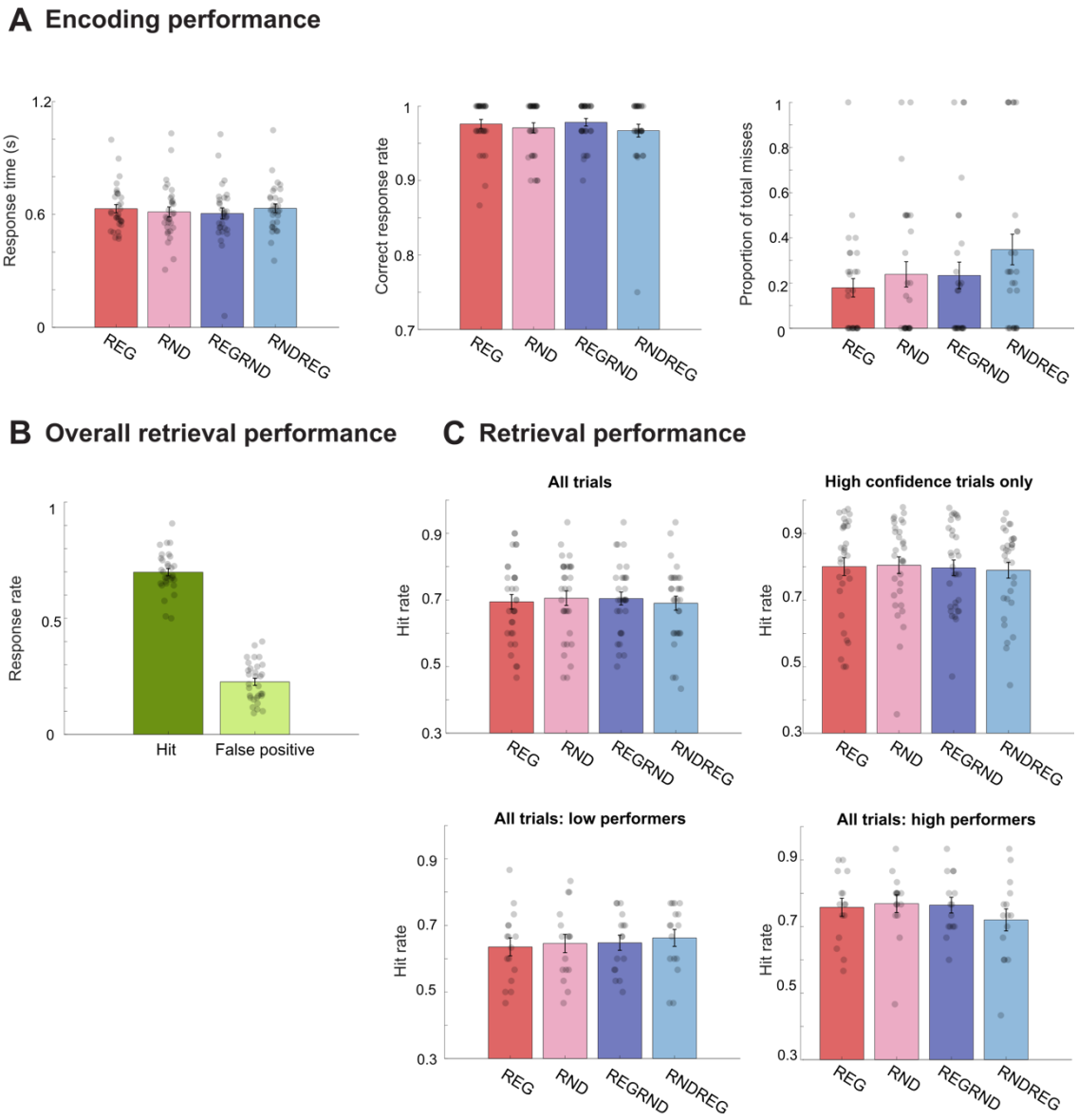


Figure 4.5 Behavioural results of the influence of the background sound on the visual task.

[A] Average response time and correct response rate for the animal vs tool judgement task during encoding session. On the right, the proportion of each condition among the missed trials was calculated for each participant and plotted. **[B]** General performance of the retrieval test. All sound conditions are collapsed. **[C]** Top: Hit rate of the retrieval test of each sound condition from all trials (left) and trials

with high confidence ratings (right). Bottom: Performance of all trials (top left) is divided into high (left) and low (right) performers based on their d' across conditions. In all figures, the error bars represent the s.e.m. Overlaid dots represent individual participants.

4.4.3 Discussion

This experiment investigated whether transitions in background sound statistics—shown to function as event boundaries in Experiment 1—could also enhance recognition performance for incidentally experienced items. Although prior research has reported improved encoding for items presented at event boundaries, such an effect was not observed in this experiment. Below, I discuss possible explanations for this discrepancy.

4.4.3.1 Potential influence of the task structure on boundary formation

One possible explanation relates to the task design. In this experiment, item recognition was tested only at the end of the experiment, and until then, participants were engaged in a different task—an animal vs. tool discrimination task. While such simple decoy tasks are commonly used (Clewett et al., 2022; DuBrow & Davachi, 2013; Dunsmoor et al., 2015; Horner et al., 2016; Kensinger et al., 2006), prior studies often include intermittent memory tests unrelated to item recognition itself (e.g., temporal order memory tests, as in Experiment 1) after each block, before testing item recognition at the end of the entire experiment (DuBrow & Davachi, 2013; Ezzyat & Davachi, 2014; McClay et al., 2023; Rouhani et al., 2020). Even when the final recognition test is unexpected, the presence of interim memory tasks may implicitly signal the importance of encoding items in context.

In the present study, the primary task did not require participants to form sequential memory traces or integrate items into coherent episodes, making it more likely that items were processed individually. Supporting this interpretation, DuBrow and Davachi (2013) conducted a study manipulating

encoding task structure to either encourage associative learning or emphasize individual item processing. They found that the memory boundary effect observed under associative learning conditions disappeared when participants were encouraged to encode items independently. This suggests that the emergence of boundary-related memory effects depends on processing strategies that promote temporal association and episodic structure.

4.4.3.2 Insufficient salience of boundary cues may limit memory enhancement

An alternative explanation is that the boundary signal in the present experiment may not have been salient enough to elicit a memory enhancement effect. The memory boost targeted here was not necessarily specific to boundary items alone; rather, similar effects have been reported in response to a wide range of salient events, including those involving threat, reward, or emotional significance (Dunsmoor et al., 2015, 2018; Greve et al., 2017; Kalbe & Schwabe, 2020; Murty & Adcock, 2014; Rouhani et al., 2018). This suggests that the memory enhancements observed in previous studies may not reflect a necessary consequence of boundary detection per se, but instead occur when event boundaries are sufficiently salient to also trigger general memory-enhancing mechanisms. In other words, prior studies that reported memory benefits at boundaries may have involved boundary cues that were inherently attention-capturing or emotionally charged (McClay et al., 2023; Rouhani et al., 2020). Supporting this, some studies using more subtle or low-salience boundary cues have similarly failed to observe enhanced item memory (DuBrow & Davachi, 2013; Horner et al., 2016; see also Dunsmoor et al., 2018).

4.4.3.3 Global regularity may undermine boundary salience

Another possible explanation for the absence of memory enhancement at transition points concerns the global structure of the stimulus. In Experiment 1—where event boundary effects emerged—stimuli consisted of five auditory ‘events’, each comprising a REG or RND sequence lasting 30–50 seconds. Furthermore, half of the blocks included a switch in pattern type (e.g., REG-

REG-REG instead of REG-RND-REG) to minimise the predictability of REG-RND transitions while still providing enough analysable trials. These features likely enhanced the perceptual salience and contextual informativeness of the transition.

In contrast, the stimuli in the current experiment featured much more frequent alternations between REG and RND within each trial (60 REG–RND transition pairs per trial) in order to present a sufficient number of test items for the final memory task. Each REG and RND segment was much shorter (3.5–5 seconds), and although I attempted to reduce predictability by occasionally presenting REGx-REGy patterns (see Methods), participants were nonetheless exposed to a highly repetitive structure. This may have allowed them to anticipate the REG–RND alternation pattern, thereby diminishing the salience and cognitive impact of each individual transition. This interpretation aligns with prior work showing that repeated exposure to a salient but behaviourally irrelevant cue can lead to habituation and reduced neural responsiveness (Sara, 2009; Sara & Bouret, 2012).

More critically, this consistent alternation may have created a higher-order regularity, leading participants to interpret each REG–RND pair as a unified unit rather than two distinct segments separated by a boundary. In effect, the numerous transitions may have rendered the stimulus highly stable at a conceptual level, with REG–RND functioning as a new “superordinate” pattern. Notably, while each REG and RND segment was uniquely generated, the possibility of abstracting such higher-order structure cannot be ruled out.

A potential future direction would be to increase the heterogeneity of the underlying regularities—for example, by introducing variation across additional auditory dimensions such as timbre, tempo, or spatial location. This would prevent transitions from being reduced to a predictable binary switch and may enhance the perception of contextual shifts.

It would also be interesting to examine whether violations of this high-level REG–RND pattern could serve as event boundaries. In this experiment, approximately 10% of the transitions were “violations”, where a REG sequence was followed by a different REG sequence (REGx–REGy), breaking the expected alternation. However, these trials were not analysed due to the absence of associated visual items. It remains an open question whether such prediction violations—despite preserving the overarching REGRND structure (which resumes immediately after the brief deviation)—are sufficient to elicit boundary-related memory effects, or whether a more fundamental contextual shift is necessary. I return to this issue in the General discussion.

In sum, while enhanced item encoding was not observed at transition points, this does not rule out the possibility that task-irrelevant background sounds can induce event boundaries. Instead, the present findings highlight the importance of task structure, stimulus salience, and contextual variability in supporting such effects.

4.5 General discussion

This study examined whether changes in the statistics of background, task-irrelevant sound sequences can induce event segmentation and influence memory for concurrently presented visual items. Across two experiments, I tested the hypothesis that transitions between regular (REG) and random (RND) tone sequences—specifically, abrupt changes from predictable to unpredictable patterns—could act as event boundaries. In Experiment 1, I found that such transitions impaired participants’ temporal order memory for items spanning the transition, suggesting that changes in auditory regularity were sufficient to segment ongoing experience into discrete episodes, even when the sound was not behaviourally relevant. In contrast, Experiment 2 tested whether these transitions would also enhance recognition memory for boundary-adjacent items but failed to find evidence for such enhancement. Together,

these findings suggest that while background auditory changes can drive event segmentation, they may not do so spontaneously. Instead, such changes may function more as contextual cues—becoming meaningful when the brain is actively engaged in forming event representations.

4.5.1 The role of surprise and context change in memory boundary formation

In this paradigm, the change in background sound patterns (e.g., REG to RND) may act both as a surprising event that violates expectations (e.g., REG violation) and as a signal that the overall structure of the sound environment has changed (e.g., switch to RND). This raises a critical question: is it the prediction error that drives event boundary formation, or is it the underlying context shift itself? This distinction is not trivial, as many paradigms investigating event boundaries inherently introduce prediction errors when a context change occurs (Clewett et al., 2020, 2025; DuBrow & Davachi, 2013; McClay et al., 2023; Pu et al., 2022; Rouhani et al., 2020).

Siefke et al. (2019) directly addressed this issue by dissociating prediction error from context change. In their study, words were presented with background colours that signalled different contextual states. In the baseline condition, colour changes were infrequent, making them surprising and enhancing memory for the association of colour-word pairs —a form of associative memory boost often observed at event boundaries (though not assessed in this chapter). However, in a critical manipulation, they reversed the statistics such that colour changes became frequent and no-change trials became surprising. If prediction error were the driving force behind the memory enhancements, memory benefits should have shifted to the no-change items. However, they did not—suggesting that it was the context shift, not the prediction error, that primarily drove the memory boundary effect. This supports the idea that changes in contextual features, rather than surprise alone, are key determinants of discontinuities in mental representations.

Supporting this, other studies have also demonstrated event segmentation in the absence of surprise (Ezzyat & Clements, 2024; Pettijohn & Radvansky, 2016; Schapiro et al., 2013; Sherman et al., 2023). For example, Schapiro et al. (2013) showed that people segment continuous experiences based on learned temporal community structure—clusters of stimuli with dense internal transitions and sparser transitions between clusters. Importantly, transitional probabilities were held constant across the nodes, ruling out surprise as a factor. Participants were still able to identify event boundaries after exposure.

Recent work by Clewett et al. (2025) adds nuance to this view by examining the role of locus coeruleus (LC) activity in memory segmentation. They found that LC activation predicted memory separation specifically at event boundaries, but not during trials occurring within a stable context. Furthermore, Wang and Egner (2023) showed that simple target detection, which is known to induce pupil-linked LC-NE activity (Swallow et al., 2019), did not impair temporal order memory. These findings support the idea that salience (e.g., surprising events) alone is insufficient to induce event boundary effects. Instead, memory segmentation appears to depend on whether an event triggers an internal update of the contextual model.

To further explore this idea, future work could investigate whether a surprising deviation in background sound alone is sufficient to trigger memory boundary formation. As discussed in Section 4.4.3.3, the stimuli used in Experiment 2 may have given rise to a higher-order regularity—a stable REG–RND alternation pattern—rendering the auditory context relatively predictable overall. Occasionally, this pattern was violated by a REG–REG transition, introducing a prediction error. However, this deviation did not signal a shift in the overarching structure of the stimulus; the regular REG–RND alternation resumed immediately afterward. In this sense, the REG–REG transition may have been perceived as noise rather than evidence of a new context. This structure parallels the REG–INT–REG stimuli discussed in Chapter 2, where the

brief interruption (INT) did not disrupt the broader regular pattern. I interpreted those findings to suggest that the brain did not treat the INT segment as a context change and thus did not reset its predictive model. By extension, if a deviation is interpreted as a transient anomaly rather than a genuine contextual shift, it may not trigger event segmentation or memory boundaries.

In Chapter 2, I also proposed that the impact of prediction error on prediction model updating depends on the inferred structure of the environment. When the global context suggests that a violation likely leads to a new pattern, the brain tends to reset its model. In contrast, when the environment implies that violations are transient and the original pattern is likely to return, the predictive model remains stable. This raises an important question: Is boundary formation similarly sensitive to these global environmental structures? Future research could directly test whether the brain flexibly distinguishes between “informative” and “noisy” deviations when deciding whether to segment experience and form episodic boundaries.

4.5.2 Broader cognitive effects of background sound

So far, this work has primarily focused on how background sound statistics contribute to event boundary formation. However, the findings also suggest that the influence of background sound extends beyond memory segmentation, affecting behaviour and perception through changes in arousal.

Arousal is closely tied to states of wakefulness, attention, stress, and motivation (Aston-Jones et al., 1999; Aston-Jones & Cohen, 2005; de Gee et al., 2024; Joshi & Gold, 2020; Waschke et al., 2019). It fluctuates dynamically in response to external events, especially those that are emotionally charged, stressful, or unpredictable (Bradley et al., 2008; Sara, 2009; Zhao, Wai Yum, et al., 2019). Several studies have demonstrated that arousal can facilitate sensory processing and behavioural responsiveness. For instance, Garrido et al. (2013) found that participants responded faster to changes in a fixation cross’s luminance when occasional pattern violations occurred in concurrently

presented background sounds. Similar enhancements in visual sensitivity have been observed when salient auditory stimuli (Ngo & Spence, 2010; Stein et al., 1996; Vroomen & Gelder, 2000) or emotionally arousing stimuli (Dahl et al., 2020; Laretzaki et al., 2010; Padmala & Pessoa, 2008; Phelps et al., 2006) coincided with visual tasks.

Consistent with these findings, I observed possible arousal-related modulation of task performance in Experiment 2. Transitions between sound types introduced violations of established sequence and were expected to be more distracting than stable sequences. However, only **RNDREG** transitions were associated with an increase in missed trials during the encoding phase, whereas **REGRND** transitions did not produce such impairments. This asymmetry may reflect a phasic arousal response specifically triggered by REG-to-RND transitions, as suggested by Zhao, Chait, et al. (2019), potentially leading to a transient boost in perceptual or attentional responsiveness.

At the same time, heightened arousal elicited by task-irrelevant stimuli can also impair performance on the primary task by diverting attention (Dahl et al., 2022; Sara, 2009; Sara & Bouret, 2012). Supporting this, I observed slower encoding responses in Experiment 1 under RND background sound compared to REG, aligning with elevated skin conductance levels during RND periods—suggesting tonic arousal increases that may have impaired performance through sustained distraction. Importantly, however, it remains unclear whether this effect reflects an impairment caused by the RND condition or, alternatively, a performance enhancement under the REG condition.

Taken together, these findings indicate that the background sound stream does more than just segment experience into discrete events. Its predictability modulates arousal in a complex way that influences ongoing cognition, attention, and memory encoding.

In sum, this work extends prior research by demonstrating that even task-irrelevant, low-level changes in sound statistics can shape the encoding of

concurrently presented visual information. This suggests that the brain actively monitors the sensory environment—even when it is not directly relevant to the task—to extract structural cues and guide cognitive processing. These insights also have promising implications for applied settings. In real-world environments, background sound may act as a subtle but powerful cue to guide attention and memory. For example, in educational contexts, strategically structured auditory environments could support event segmentation and improve long-term retention. Prior research has shown that individuals who are better at perceiving and segmenting events tend to have enhanced memory for the overall episodes, even weeks later (Flores et al., 2017; Gold et al., 2017; Sargent et al., 2013). Leveraging multimodal cues—such as synchronising auditory and visual shifts—may enhance boundary perception and facilitate memory organisation.

5. Chapter 5: General Discussion

5.1 Summary of findings

Everyday auditory environments are rich and dynamic, shaped by constantly changing acoustic patterns. By investigating how the brain processes regularities in these complex contexts, I have taken steps toward advancing our understanding of auditory processing—moving beyond the insights gained from studies that rely primarily on simple stimuli.

Chapter 2 investigated how the brain utilises prior experience to guide ongoing processing of sound sequences. Comparison of EEG data with computational models employing various training window sizes revealed that even during passive listening, the auditory system flexibly evaluates the relevance of past information. When the experimental context indicated a stable environment, the brain retained and integrated past context, whereas it minimised the influence of prior experience when the environment suggested volatility. These findings suggest that even through passive exposure, the brain tracks environmental stability and dynamically adjusts its predictive strategies accordingly—revealing an adaptive mechanism that supports efficient perception in ever-changing environments.

Chapter 3 examined how the predictability of a preceding auditory sequence influences the efficiency of detecting regularities in a subsequent sound stream. The study revealed that prior exposure to a deterministic context delayed the emergence of the neural signature associated with discovering a new regular pattern. Notably, the observed EEG dynamics were not fully captured by two commonly used computational models—IDyOM and D-REX. This discrepancy highlights a gap between current model predictions and actual brain responses, casting doubt on the notion that the brain operates according to Bayesian principles.

Finally, Chapter 4 explored whether changes in the statistical structure of background sounds—despite being behaviourally irrelevant—could act as event

boundaries and influence the encoding of concurrently presented visual information. Results from two experiments provided partial support for this hypothesis. While the effects were weaker than those typically observed with task-relevant boundaries, the findings suggest that statistical changes in the auditory background can influence memory organisation across modalities, highlighting the subtle yet pervasive impact of unattended sound on broader cognitive functions. Additionally, skin conductance measurements revealed that the predictability of background sounds modulated arousal levels, offering further evidence that auditory regularities can influence the brain's global state, even when outside the focus of attention.

Overall, this thesis advances our understanding of how the brain processes complex, dynamically changing auditory regularities, and how this processing shapes broader neural and cognitive functions. The findings underscore the importance of studying auditory perception within the rich, ever-changing contexts that characterise real-world listening. Moreover, this series of experiments highlights the powerful role of background sound statistics, revealing that the brain treats such information as a meaningful cue for inferring the state of the environment—even when it holds no explicit behavioural relevance.

5.2 Implications

This thesis provides compelling evidence that the brain does more than passively detect the acoustic features of background, task-irrelevant sounds—it actively interacts with them, extracting structure and drawing inferences that extend far beyond simple sensory analysis.

When considering the potential influence of unattended auditory input, several levels of engagement can be proposed. At the most basic level, the brain might entirely ignore such input. A step beyond this would involve passive, bottom-up processing that registers the signal without higher-order

interpretation. However, prior research indicates that the brain often goes further, engaging in predictive inference even when sounds are not the focus of attention. As reviewed in Chapter 1, phenomena such as mismatch negativity (MMN) demonstrate that the brain forms expectations about incoming auditory input based on prior experience, even in the absence of directed attention (Bendixen et al., 2007; Bendixen & Schröger, 2008; Tivadar et al., 2021). Similarly, the alignment between sustained neural responses and predictions generated by ideal observer models—which predict future events based on past experiences—suggests that the brain actively constructs internal representations of auditory sequences by tracking and exploiting statistical regularities over time (Barascud et al., 2016; Bianco et al., 2025; Hu et al., 2024; Zhao et al., 2025).

The formation of expectations enables us to detect change points in the sensory environment, often signalled by violations of predictions. Such change point detections indicate that internal models based on past input are no longer reliable, prompting an update of the temporal reference window used for generating predictions. Previous studies have shown that the brain is highly sensitive to abrupt changes in the statistical structure of auditory input, responding with a sharp reduction in sustained neural activity (Barascud et al., 2016; Bianco et al., 2025; Zhao et al., 2025) and activation of the pupil-linked locus coeruleus–norepinephrine (LC-NE) system (Basgol et al., 2025; Zhao, Chait, et al., 2019). However, the downstream consequences of such an abrupt change point detections for neural processing have remained less well understood. This thesis contributes to filling that gap by elucidating how change point detection impacts ongoing neural computations.

In Chapter 2, I showed that upon detecting a change, the brain shortens its temporal reference window for prediction (i.e., run length). In Chapter 3, I demonstrated that the detection of the abrupt change paradoxically delayed the discovery of new regularities in the environment. In Chapter 4, I found that

change point detections modulated the formation of visual memory, suggesting that its influence extends beyond the auditory domain.

Importantly, change point detection is not exclusive to the auditory domain. For instance, it is well documented in the domain of sequential decision-making, where extensive research has shown that individuals adjust their predictive models in response to inferred changes in environmental statistics (Glaze et al., 2015; McGuire et al., 2014; Nassar et al., 2010, 2012). Similarly, as discussed in Chapter 4, memory research indicates that contextual shifts act as event boundaries, segmenting experience into discrete episodes and shaping the organisation and retrieval of long-term memory (Clewett et al., 2020, 2025; DuBrow & Davachi, 2013; Ezzyat & Davachi, 2011; Horner et al., 2016). The fact that change-point sensitivity emerges in auditory perception, decision-making, and memory suggests that it may reflect a domain-general computational principle—one that enables the brain to remain attuned to environmental dynamics by flexibly adjusting internal models across both time and modality. Taken together, the findings support the idea that change point detection is a core component of the brain's adaptive architecture, supporting efficient and flexible behaviour in an unpredictable world.

Furthermore, this thesis proposes an additional layer of auditory processing beyond simple prediction formation: the inference of broader environmental structure from the evolving statistics of ongoing sound sequences. In Chapter 2, I observed that the brain does not always interpret prediction violations as evidence of environmental change. When violations occurred in a context where such deviations consistently indicated a shift in the auditory scene, the brain responded by shortening its temporal reference window for future predictions. In contrast, when the same violations occurred in a context where they did not typically signal a change, the brain continued to rely on pre-violation information. These differential responses to identical violations suggest that the brain learns about the stability of the environment through exposure and uses this inferred stability to evaluate whether a given

violation is more likely to reflect a genuine change or mere noise. In volatile environments, violations are more likely to signal change, whereas in stable contexts, they are more likely treated as random fluctuations.

Such sensitivity to environmental volatility has been widely studied in the decision-making literature (Behrens et al., 2007; Glaze et al., 2015; Piray & Daw, 2024). However, Chapter 2 provided compelling evidence that the similar computational principles operate on a much faster timescale. These findings suggest that the brain continuously and automatically monitors environmental dynamics, flexibly adjusting its internal models to optimise perception and behaviour—even in passive, task-irrelevant listening contexts.

Although further work is needed to fully validate this claim, the present findings represent a promising step toward understanding how the brain dynamically adapts to changes in its environment. They raise questions about whether the computational strategy of change detection and adaptive model updating is shared across cognitive domains and timescales. If so, auditory regularity tracking may offer a powerful window into the brain's core computational architecture. This possibility suggests that neural responses to auditory regularities may serve as a broader index of the brain's inferential processes—extending beyond sound processing. The auditory sequences used in this thesis were devoid of semantic or emotional content, providing a relatively unbiased platform for examining fundamental neural computations. In turn, these paradigms could generate hypotheses and models relevant to other domains, including visual perception, decision making, and memory.

5.3 Limitations

While this thesis aimed to bridge the gap between experiments using simplified auditory stimuli and the complexity of real-world soundscapes, the stimuli employed remained artificial. This design choice was deliberate,

enabling precise testing of hypotheses concerning dynamic tracking of auditory regularities while minimising potential confounds. All sound sequences were constructed from pure tone pips drawn from a fixed set of 20 frequencies—chosen to avoid unintentional salience or emotional associations that might arise from broader frequency distributions or more naturalistic sounds. Furthermore, the auditory patterns used in this thesis were either fully random or fully deterministic. While such extremes are rare in real-world listening—where statistical structure is typically more graded and probabilistic—this binary manipulation offered an advantage of providing statistical power, which was particularly important given the inherent noisiness of the EEG data. Future research could build on these findings by incorporating more ecologically valid, naturalistic auditory patterns to assess whether the observed effects generalise beyond the controlled conditions used here. Such work would help extend and validate the mechanisms identified in this thesis within more complex, real-world listening environments.

In terms of methodology, the use of EEG enabled high temporal resolution and allowed us to track rapid changes in neural responses over time. However, this came at the cost of spatial resolution, limiting our ability to precisely localise the neural generators of the observed signals. Moreover, the sustained neural responses that formed a central focus of this work are particularly susceptible to low-frequency noise and slow drift—sources of artefact that overlap with the frequency range of the signal itself. As a result, conventional denoising techniques such as high-pass filtering were not viable without risking the loss of the signal of interest. To mitigate this, analyses relied on trial-averaging to reveal consistent response patterns. However, this approach precludes trial-level analyses, which are essential for understanding how the brain incrementally learns and adapts to statistical structure over the course of an experiment.

Finally, the computational models used in this thesis served to formalise task-specific hypotheses and guide the interpretation of empirical findings,

rather than to directly fit the neural data. For example, in Chapter 2, models were implemented using default parameters to explore how memory span might shape auditory scene representation, without aiming to identify the best-fitting model at the individual or group level. While this approach provided valuable conceptual insights, future work could benefit from adopting a more systematic model-fitting framework to more precisely characterise the computational mechanisms underlying regularity processing in dynamic auditory environments.

5.4 Future direction

While the present findings offer valuable insights into the passive listening brain's capacity to track changes in background auditory statistics, several critical questions remain for future research.

First, although this thesis raised the possibility that the passive brain can infer environmental volatility and adjust its prediction strategy accordingly, this conclusion remains tentative. In Chapter 2, I compared two types of auditory sequences: one in which deviations consistently signalled a structural change, and another in which deviations does not always lead to a structural change. While this design offers initial support for volatility-sensitive inference, it remains unclear whether the adjustment of predictive strategy occurs on a trial-by-trial basis or emerges gradually as the brain accumulates evidence about the environment over time. Furthermore, real-world environments rarely operate on simple binary rules used in the experimental paradigm; instead, they often involve more complex, fluctuating patterns with varying rates of change. Future studies should investigate how much statistical complexity the passive listening brain can accommodate and how this capacity is shaped by the duration of exposure.

For example, what happens when genuine structural changes are embedded within random fluctuations? Recent findings in the decision-making

literature suggest that, even in such ambiguous contexts, the brain can distinguish volatility from stochasticity—when active attention is engaged, possibly by tracking features such as autocorrelation across successive events (Piray & Daw, 2024). However, it remains unclear whether, and how, the passive brain—without explicit task demands—can make similar distinctions. Comparing what is already known about inference under active engagement with the responses observed in passive contexts could clarify whether rapid auditory tracking relies on the same mechanisms as higher-order decision-making—or whether it reflects modality-specific heuristics optimised for fast sensory environments.

Understanding how the brain samples and interprets environmental information is not only of theoretical interest—it also holds significant implications for mental health and neuropsychiatric disorders. Maladaptive inferences about environmental volatility have been implicated in conditions such as psychosis, anxiety, and autism (Browning et al., 2015; de Berker et al., 2016; Lawson et al., 2014, 2017; Powers et al., 2017). While these patterns have primarily been observed in the context of sequential decision-making tasks, the findings presented in this thesis suggest that similar maladaptive mechanisms may also be reflected in how individuals track and respond to changes and volatility in the background auditory sequences.

Investigating this possibility could provide valuable insights. First, it may deepen our understanding of how neurodivergent individuals experience and interpret auditory signals. Second, and more critically, it raises the potential for identifying sustained neural responses to emotionally neutral, semantically stripped sound sequences as candidate markers for clinical symptoms—or even as therapeutic targets. Future work could explore whether stable auditory input might help recalibrate maladaptive beliefs about environmental uncertainty and volatility, akin to extinction-based strategies used in cognitive behavioural therapy. For instance, pairing perceived volatility with a consistently stable

sound sequence may facilitate more accurate belief updating in individuals prone to anxiety or hypervigilance.

A related avenue for future research concerns the broader influence of background sound statistics on perception and behaviour. As demonstrated in Chapter 4, changes in the statistical structure of background sounds can modulate processing in other sensory modalities, indicating that the auditory environment shapes how we engage with the world more broadly. This raises the possibility that environmental auditory cues could be strategically leveraged in applied settings. For instance, incorporating structured auditory regularities into the design of public spaces or workplaces may promote more efficient navigation, improve focus, and mitigate cognitive load in complex real-world settings. While there is a growing body of research on the impact of soundscapes and ambient noise on cognition and behaviour (Angel et al., 2010; Baijot et al., 2016; Haake, 2011; Woods et al., 2024), the findings from this thesis contribute to this line of work by emphasising the potential relevance of statistical structure as a factor influencing perceptual and cognitive processing.

6. Appendix Chapter: Is Speaker Size a Salient Auditory Feature?

6.1 Summary

This chapter presents a set of supplementary experiments conducted independently of the core research question addressed in the main thesis. As reviewed in Chapter 1, salient sounds often convey biologically relevant information and capture attention through bottom-up mechanisms. Here, I examined whether vocal cues related to speaker size can serve as auditory salience signals. In Experiment 1, an online behavioural study demonstrated that listeners are highly proficient at discriminating vowels produced by speakers of different sizes. In Experiment 2, the same stimuli were presented in a controlled laboratory environment while ocular dynamics—specifically, microsaccadic inhibition (MSI) and the pupil dilation response (PDR), both considered objective indices of auditory salience—were measured. Contrary to our hypothesis, voices perceived as larger in size did not elicit stronger ocular responses. These findings suggest that vocal size may not constitute a primary acoustic feature driving bottom-up auditory salience.

6.2 Introduction

The ability to infer body size from vocal cues is a fundamental skill in social communication across the animal kingdom. This capacity serves crucial functions, such as selecting mates, evaluating potential rivals, and maintaining territorial boundaries. A broad range of species—including deer (Reby et al., 2005), frogs (Fairchild, 1981), monkeys (Ghazanfar et al., 2007), and humans (Ives et al., 2005; D. R. R. Smith et al., 2005; D. R. R. Smith & Patterson, 2005)—share this sensitivity to acoustic indicators of body size, highlighting its evolutionary importance.

In human speech, cues to body size are primarily conveyed through two physiological features: vocal tract length (VTL) and glottal pulse rate (GPR). VTL is closely linked to overall body size and shows a near-linear inverse relationship with vowel formant frequencies—longer vocal tracts result in lower formant frequencies (Fitch & Giedd, 1999). In contrast, GPR refers to the rate at which the vocal folds open and close, which largely determines the perceived pitch of the voice. It is shaped by the length and mass of the vocal folds, both of which typically increase with development (Titze, 1989).

The influence of VTL and GPR on perceived speaker size has been extensively studied using precise acoustic manipulations (Ghazanfar et al., 2007; Ives et al., 2005; D. R. R. Smith et al., 2005; D. R. R. Smith & Patterson, 2005; von Kriegstein et al., 2006, 2007). A widely used tool in this research is the STRAIGHT vocoder (Kawahara et al., 1999; Kawahara & Irino, 2004), which allows the separation of a vowel's GPR contour from its spectral envelope. This capability enables flexible resynthesis of vowels with independent control over GPR and VTL—for example, converting an adult male voice into one resembling a child's.

Using STRAIGHT, D.R.R. Smith et al. (2005) showed that listeners could detect differences in VTL as small as 6–10% in a two-alternative forced choice task. Given that the just-noticeable difference for loudness is around 10% (Miller, 1947), this highlights the fine-grained sensitivity of human listeners to vocal tract cues. D.R.R. Smith and Patterson (2005) further demonstrated that listeners could estimate a speaker's size from a single voice token, without comparison stimuli. Their results confirmed that while both GPR and VTL contribute to size perception, VTL plays the more dominant role. A meta-analysis by Pisanski et al. (2014) reinforced this conclusion.

Importantly, this sensitivity to vocal size cues is not unique to humans; Ghazanfar et al. (2007) extended these findings to rhesus monkeys. In their study, untrained monkeys were presented with 'coo' calls that had been modified using STRAIGHT to simulate different vocal tract lengths. The

monkeys looked at larger faces when hearing larger-sounding coos and at smaller faces when hearing smaller-sounding coos—indicating a cross-species sensitivity to vocal cues of size.

Given the biological importance of vocal size cues and their demonstrated cross-species relevance, it is plausible that this information may function as salient auditory features that can automatically engage perceptual systems through bottom-up processes. Just as acoustic properties like loudness and roughness have been shown to increase perceptual salience (Arnal et al., 2015; N. Huang & Elhilali, 2017; Liao et al., 2016; Zhao, Wai Yum, et al., 2019), voices that convey larger body size may likewise have a stronger capacity to draw attention in a bottom-up manner.

What makes a sound salient remains poorly understood. In the visual domain, salience has been effectively studied using eye movements as an index of bottom-up attentional capture (Krauzlis et al., 2019; Parkhurst et al., 2002; R. J. Peters et al., 2005; Veale et al., 2017). However, the auditory domain lacks a similarly well-established method for measuring stimulus-driven attentional captures, making it hard to test auditory salience systematically. Recent research has suggested that ocular dynamics—such as microsaccades and pupil dilation—may offer an objective window into auditory salience (Liao et al., 2016; Zhao, Wai Yum, et al., 2019). Nevertheless, only a limited set of acoustic features has been validated to modulate these ocular responses, leaving open the question of whether other perceptually relevant features, such as speaker size, can also influence them.

Thus, this study aims to broaden our understanding of auditory salience by investigating an additional potentially salient auditory feature. This chapter explores whether vocal size cues contribute to auditory salience using synthetic vowel stimuli that vary in perceived speaker size. Experiment 1 validated listeners' sensitivity to these size manipulations in a large-scale online study and Experiment 2 assessed the salience of these cues by measuring oculomotor responses in a controlled laboratory setting.

6.3 Experiment 1

Before conducting the main eye-tracking experiment, the stimulus set was validated through an online study. This study aimed to confirm that participants could accurately differentiate size differences conveyed by the vowel sounds.

6.3.1 Methods

6.3.1.1 Stimuli

Three vowels (/a/, /e/, /i/) were recorded from a female speaker with a height of 155 cm and subsequently manipulated using WORLD (Morise et al., 2016), a vocoder similar to STRAIGHT but optimised for lower computational cost. The vocal tract length (VTL) parameter was systematically adjusted to simulate speakers of varying heights—142 cm, 155 cm, 169 cm, 184 cm, and 201 cm—based on values reported in D.R.R. Smith and Patterson (2005). The glottal pulse rate (GPR) was fixed at 71% of the original to minimise variation in roughness across stimuli, a known contributor to auditory salience (Arnal et al., 2015; Zhao, Wai Yum, et al., 2019). All stimuli were 700 ms in duration and root-mean-square (RMS) normalised.

6.3.1.2 Procedure

The experiment was designed and hosted on the Gorilla platform (www.gorilla.sc; Anwyl-Irvine et al., 2020). Participants were asked to make relative size judgments based on pairs of vowel sounds, presented with a 500 ms inter-stimulus interval. From a pool of 105 possible sound pairs, 34 were randomly selected for each participant. To control for potential order effects, a second set of 105 reversed-order pairs was generated, and half of the participants were randomly assigned to this reversed pool. After hearing each

pair, participants indicated which sound they believed was produced by a larger individual.

To ensure participants were attentive during the online experiment, six catch trials were randomly interspersed throughout the task. In these trials, one of the vowel sounds contained a brief 50 ms silent gap (a “blip”), and participants were asked to identify which sound contained it. The blip was equally likely to occur in the first or second sound. Participants who failed more than one catch trial were classified as outliers and excluded from further analysis (four participants were excluded).

Prior to the main experiment, participants completed a practice session for the blip detection task, and only those who correctly identified all blips were allowed to continue; five participants were excluded at this stage. For the size judgement task, no practice trials were provided to avoid biasing participants. However, participants had the opportunity to listen to example stimuli before starting the main experiment.

6.3.1.3 Participants

Participants aged between 18 and 40 years with no reported hearing problems were recruited via Prolific (www.prolific.co). All participants were required to use a laptop or desktop computer and wear headphones. A headphone screening test (Milne, Bianco, et al., 2021) was administered to verify headphone use, resulting in the exclusion of 62 participants. An additional 9 participants were excluded: 5 for failing the practice session and 4 for failing more than one catch trial (see above). The final sample comprised 140 participants (83 female; mean age = 30.54 ± 6.26). All participants were naïve to the purpose of the study and were instructed to complete the experiment in a quiet environment while seated comfortably. All experimental procedures were approved by the research ethics committee of University College London.

6.3.2 Results

6.3.2.1 Consistent size ranking was observed across vowels

Figure *6.1A* illustrates the number of presentations for each sound across all participants, showing that each sound was presented approximately 650 times. To determine the size ranking of the 15 sounds, the proportion of trials in which each sound was judged as larger than its paired counterpart was calculated relative to its total number of presentations. As shown in **Figure *6.1B***, sounds representing larger sizes were more frequently judged as larger than their paired sounds. This size ranking was consistently observed across all vowel categories (**Figure *6.1C***), and the order remained stable even when averaged across vowels (**Figure *6.1D***).

Importantly, the potential effect of presentation order on participants' judgments was evaluated by comparing the number of "larger" responses attributed to the first versus the second sound. This analysis found no significant order bias (paired t-test; $t(139) = -0.37$, $p = 0.71$).

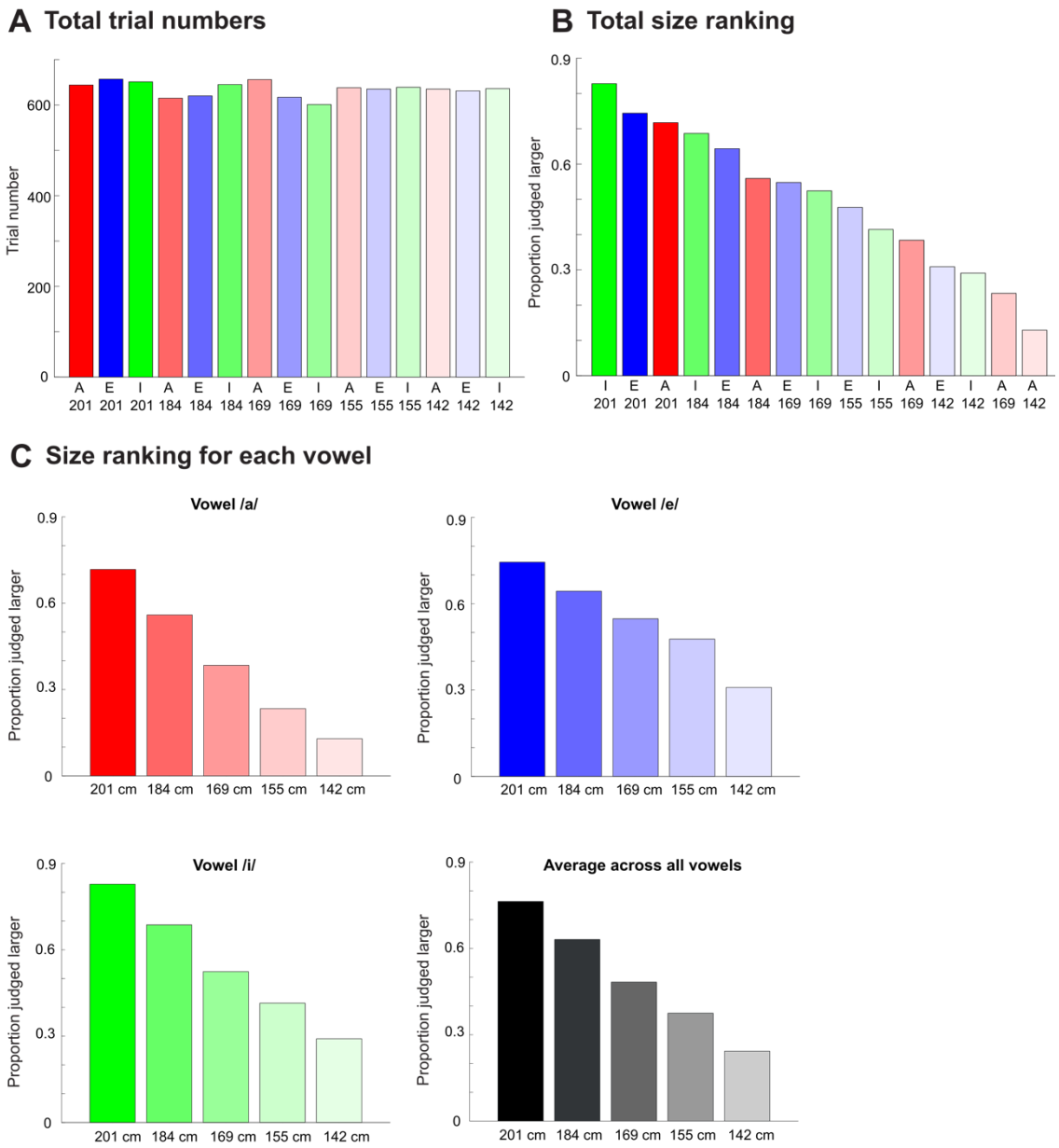


Figure 6.1 Results of the subjective size judgement.

[A] Total number of trials each sound was presented, summed across all participants. Colours indicate vowel conditions (green: /i/, blue: /e/, red: /a/), with fainter colours representing sounds associated with smaller sizes. **[B]** Proportion of trials in which each sound was judged as larger than its paired counterpart. Colours indicate vowel conditions (green: /i/, blue: /e/, red: /a/), with fainter colours representing sounds

associated with smaller sizes. **[C]** Same data as [B], reorganised by vowel condition to show size judgement within each vowel category. The bottom-right plot displays the average across the three vowel conditions.

6.3.2.2 Identifying ambiguous sound pairs in size judgement

To assess which sound pairs participants found difficult to discriminate, I analysed their size judgments across all pairwise combinations. For each sound condition (columns in **Figure *6.2A***), I calculated the proportion of trials in which it was judged as larger than its paired sound (rows in **Figure *6.2A***), by dividing the number of “larger” responses by the total number of presentations for that specific pair. Diagonal cells were excluded (set to NaN), as identical sounds were never presented together.

Figure *6.2B* illustrates the ideal response pattern, where the sound representing the larger size is always judged as larger, and performance for same-size pairs remains at chance. Overall, participants’ responses broadly aligned with this pattern: brighter cells—indicating higher proportions of “larger” responses—clustered in the lower-left triangle.

Nevertheless, some sound pairs showed evidence of confusion. In particular, pairs involving adjacent size steps (e.g., 169 cm vs. 155 cm) were more difficult to distinguish reliably. The mid-range 169 cm stimuli were especially prone to misjudgement, suggesting that this condition may lie near a perceptual boundary within the stimulus continuum.

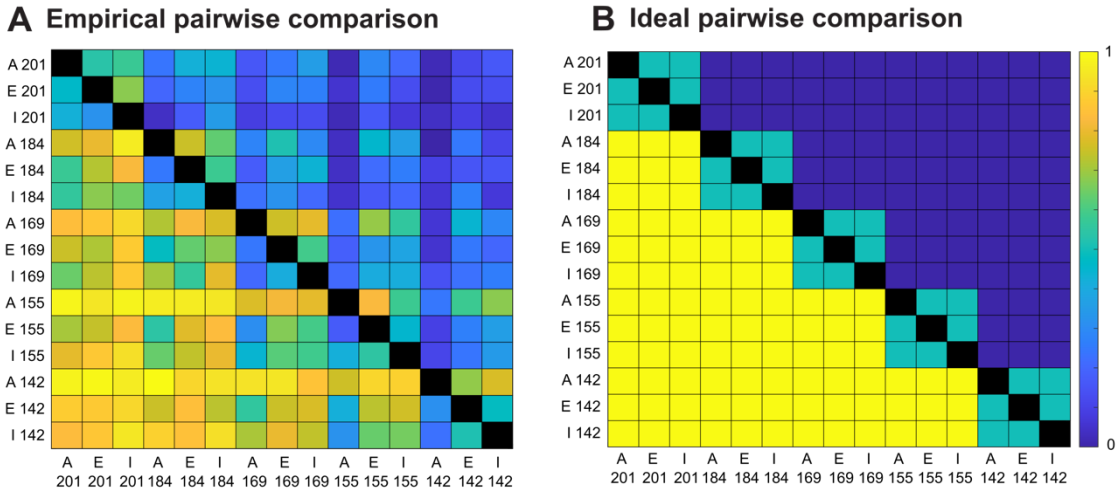


Figure 6.2 Judgement matrix for each sound pair.

[A] Proportion of trials in which the sound in each column is judged as larger than its paired sound (row). Brighter colours indicate higher proportions. Diagonal cells are set to NaN. **[B]** Ideal response matrix assuming perfect discrimination.

6.3.3 Discussion

This study demonstrated that listeners can reliably judge differences in speaker body size based on vowel sounds. Across vowel types, participants consistently perceived voices with longer VTL as belonging to larger speakers, producing a size ranking that closely matched the intended size manipulation. This robust performance is particularly noteworthy given the study's online format, which involved naïve participants using varied hardware and listening environments.

Given the consistency of the size ranking across vowels, subsequent eye-tracking analyses treated vowel identity as a controlled factor and focused solely on differences in size. Additionally, pairwise analysis (**Figure *6.2***) revealed that the mid-range stimulus (169 cm) was particularly prone to

confusion with adjacent size conditions. To avoid potential ambiguity, this stimulus will be excluded from the following eye-tracking experiment.

Crucially, this experiment also validated the robustness of size perception from voice in a broader and more ecologically valid context. Previous studies in this area (Ghazanfar et al., 2007; Ives et al., 2005; D. R. R. Smith et al., 2005; D. R. R. Smith & Patterson, 2005; von Kriegstein et al., 2006, 2007) relied on small samples (typically fewer than 15 participants) and were conducted under controlled laboratory conditions. While these foundational studies demonstrated the perceptual relevance of vocal size cues, they left open the question of whether this ability generalises across populations and environments.

By testing a larger and more diverse sample (n = 140) in a less controlled online environment, the present study provides compelling evidence that vocal size discrimination is a robust and generalisable perceptual skill—likely reflecting a fundamental aspect of human auditory processing. Notably, participants succeeded in the task with minimal instruction and without any practice trials. Despite considerable variability in listening conditions—including background noise, device quality, and participant posture—listeners consistently detected size-related vocal cues. This finding suggests that the perceptual features underlying speaker size estimation are highly salient and readily accessible, resilient to individual and environmental variability.

6.4 Experiment 2

This chapter investigates whether perceived speaker size, conveyed through vowel sounds, serves as a salient auditory cue that automatically attract attention in a bottom-up manner. To examine this, oculomotor responses were recorded from participants passively exposed to the same stimuli used in Experiment 1. Specifically, the analyses focused on two measures: microsaccadic inhibition (MSI) and the pupil dilation response (PDR).

Microsaccades (MS) are small, involuntary eye movements that occur roughly once or twice per second (Rolfs, 2009). These movements are closely linked to attentional sampling mechanisms involving the frontal eye fields and the superior colliculus (Hafed et al., 2009, 2015; Krauzlis et al., 2013; Peel et al., 2016; Rolfs, 2009; Rucci & Poletti, 2015; C.-A. Wang & Munoz, 2015; Zénon & Krauzlis, 2012). When attention is rapidly captured by a surprising events, microsaccades are temporarily suppressed (Hafed & Clark, 2002; Rolfs et al., 2008; Zhao, Wai Yum, et al., 2019) —a phenomenon known as microsaccadic inhibition (MSI). This suppression is considered to reflect a temporary interruption of the brain's spontaneous exploration of the visual environment, allowing prioritisation of processing potentially important events (Contadini-Wright et al., 2023; Zhao, Wai Yum, et al., 2019).

The degree of MSI—typically characterised by faster onset, later offset, and fewer microsaccades during the inhibition phase—has been shown to be modulated by the salience of visual stimuli (Bonneh et al., 2015; Rolfs et al., 2008). Zhao, Wai Yum, et al. (2019) extended this finding to the auditory domain, demonstrating that the subjective salience ranking of sounds correlates with the degree of MSI. This raises the possibility that MSI reflects domain-general stimulus salience and the extent of stimulus-evoked bottom-up attentional capture.

In contrast, the pupil dilation response (PDR) is linked to activity in the locus coeruleus (LC), the brain's primary source of norepinephrine (NE; Aston-Jones & Cohen, 2005; Joshi et al., 2016). Since NE modulates arousal and global vigilance by enhancing neuronal gain (Sara, 2009; Sara & Bouret, 2012), PDR serves as an indirect measure of arousal. Supporting this connection, Joshi et al. (2016) showed that loud, arousing tones evoke increases in LC activity and larger pupil dilations. Additionally, the superior colliculus (SC) has been implicated in mediating pupil responses to salient stimuli in the visual domain (C.-A. Wang et al., 2014; C.-A. Wang & Munoz, 2015). However,

findings in the auditory domain are less consistent (Liao et al., 2016; C.-A. Wang et al., 2014; Zhao, Wai Yum, et al., 2019).

Although both MSI and PDR are evoked by surprising stimuli, they exhibit distinct temporal dynamics. In **Figure *6.3***, MSI emerges around 300 ms after stimulus onset and persists for several hundred milliseconds, whereas the pupil dilation response (PDR) begins later—after 500 ms in **Figure *6.3***—and peaks around one second post-onset (see also Contadini-Wright et al., 2023). These differences suggest that MSI and PDR may reflect distinct stages of neural processing, with MSI indicating early attentional orienting and PDR indexing a later, arousal-related response. Alternatively, the temporal dissociation may arise from physiological factors—for example, delays in the neural circuit linking arousal modulation to the musculature controlling pupil size. While some studies have reported correlations between the two responses (Johnston et al., 2022; C.-A. Wang et al., 2022; C.-A. Wang & Munoz, 2021), the precise nature of their relationship remains under debate.

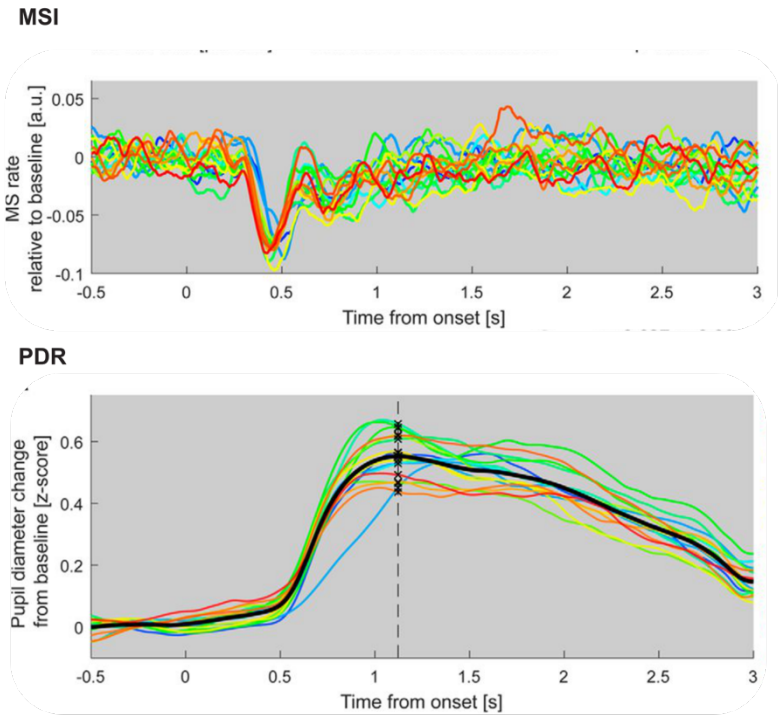


Figure 6.3 Examples of sound-evoked ocular dynamics.

Top: Microsaccadic inhibition (MSI) induced by various sound stimuli. The y-axis indicates the microsaccade rate (events per second) relative to baseline. Each coloured line represents a different sound condition. Bottom: Pupil dilation response (PDR) evoked by the same sound stimuli. Each coloured line corresponds to a different sound; the thick black line indicates the average response across conditions. The dashed line marks the peak of the average PDR. Adapted from Zhao, Wai Yum, et al. (2019).

In this experiment, I investigated whether MSI and PDR are sensitive to speaker size cues embedded in vowel sounds. I hypothesised that voices perceived as coming from larger speakers would be judged as more salient—or more threatening—than those from smaller speakers, and would thus elicit stronger MSI and/or greater PDR.

6.4.1 Methods

6.4.1.1 Stimuli

The stimuli used in this experiment were generated in the same manner as those in Experiment 1. The same three vowel sounds (/a/, /e/, /i/) were included; however, sounds corresponding to the 169 cm height condition were omitted, based on findings from Experiment 1. This resulted in a total of 12 sound stimuli used in the current experiment. Based on the result from Experiment 1, the subsequent analyses treat vowel identity as a controlled factor and focused solely on differences in size.

6.4.1.2 Procedure

Participants were seated in a dimly lit, acoustically shielded room (IAC triple-walled sound-attenuating booth), with head movements minimised using a chinrest. They passively listened to auditory stimuli delivered diotically via

headphones (Sennheiser HD558) at a comfortable listening level (adjusted by each participant). During sound presentation, participants maintained visual fixation on a central cross displayed on a monitor (24-inch BENQ XL2420T; 1920 × 1080 resolution; 60 Hz refresh rate) positioned 65 cm away.

Each participant completed five blocks of trials, with rest breaks between blocks. Each block contained 48 trials, comprising 12 distinct sound stimuli presented four times in a randomised order. Intertrial intervals were jittered between 6 and 7 seconds. Stimulus presentation was controlled using the Psychtoolbox package (Psychophysics Toolbox Version 3) in MATLAB (2018a, The MathWorks, Inc.).

6.4.1.3 Eye tracking recording

Ocular dynamics were recorded using an EyeLink 1000 Desktop Mount eye tracker (SR Research), positioned below the monitor and sampling at 1000 Hz. A standard five-point calibration procedure was performed prior to each block. Each trial commenced only after the system confirmed that the participant's eyes were open and fixated on the central fixation cross. Participants were instructed to blink naturally throughout the experiment.

6.4.1.4 Microsaccade preprocessing and analysis

To identify microsaccade (MS) events from horizontal eye movement data, the following criteria were applied, based on Zhao, Wai Yum, et al. (2019): (a) velocity exceeding six times the standard deviation within each block; (b) event duration between 3 ms and 100 ms; (c) binocular detection with onset disparity less than 10 ms between eyes; and (d) a minimum interval of 50 ms between successive MS events. The onset of each detected MS was coded as 1, while all other time points were coded as 0. The resulting binary time series was then epoched from –1 to +5 seconds relative to sound onset.

To examine MS dynamics over time, MS event series were smoothed using a causal exponential kernel (decay parameter $\alpha = 1/50$ ms). For this analysis, the four sound conditions were grouped into two categories: *large* (201

cm and 184 cm conditions) and *small* (155 cm and 142 cm conditions). For each participant, MS event data were first averaged within each group (large, small), then smoothed, and subsequently z-score normalised using the baseline window from –0.2 to 0 seconds relative to stimulus onset.

Trials with excessive blinking were excluded, as they can compromise data quality. The number of blinks per epoch was calculated, and the top 10% of blink-heavy epochs across all participants were discarded. Additionally, participants with extremely low MS rates (<0.5 events/s across trials) during the pre-stimulus (–1 to 0 s) time window were excluded from further analysis. As a result, three participants were flagged as outliers.

6.4.1.5 Pupillometry preprocessing and analysis

Only data from the left eye were analysed. Periods corresponding to blinks, partial blinks, or when gaze deviated more than 100 pixels from the fixation cross were excluded. These missing segments were reconstructed using shape-preserving piecewise cubic interpolation. The resulting time series were smoothed using a 50 ms Hanning window.

Data were then epoched from 1 second before to 5 seconds after sound onset, z-score normalised within each block, and baseline-corrected by subtracting the median pupil size in the −0.2 to 0 s window preceding stimulus onset from each trial. As in the MS analysis, sound conditions were grouped into large and small speaker size categories and averaged per condition.

Trials contaminated by excessive blinking were excluded using the same criteria as in the MS analysis. In addition, trials with average responses (0–5 s post-stimulus) falling outside ±2 standard deviations (SD) of the condition mean were discarded. Blocks with baseline variability (−0.2 to 0 s) exceeding ±2 SD of the grand average were also excluded. Furthermore, blocks in which participants maintained fixation for less than 65% of the time were removed. Data from one participant were entirely excluded, as more than half of the data were identified as outliers under these criteria.

6.4.1.6 Statistical analysis

For both MS and PDR data, the difference between sound conditions (large vs. small) was calculated for each participant. These individual differences were then subjected to bootstrap resampling (Efron & Tibshirani, 1994). A difference was considered statistically significant if more than 99% of the bootstrap iterations fell consistently above or below zero ($p < .01$).

6.4.1.7 Participants

Thirty-one paid participants aged 18 to 40 were recruited for Experiment 2. For the MS analysis, three participants were excluded due to the data quality (see above for details), resulted in a final sample of twenty-eight participants (23 females; mean age = 23.5, ± 4.1). For the PDR analysis, one participant was excluded due to the data quality (see above for details), resulted in a final sample of thirty participants (23 females; mean age = 24.1, ± 4.7). All participants reported no history of hearing or neurological disorders. All experimental procedures were approved by the research ethics committee of University College London, and written informed consent was obtained from each participant.

6.4.2 Results

6.4.2.1 Microsaccadic inhibition did not reflect speaker size

Microsaccadic events (see Methods) were extracted for each trial (**Figure *6.4A left***). A rapid reduction in MS events—microsaccadic inhibition (MSI)—was consistently observed across trials immediately following sound onset. To compare the degree of MSI between conditions in more detail, the MS event data were smoothed (see Methods) and averaged within two groups based on speaker size: *large* (201 cm and 184 cm conditions) and *small* (155 cm and 142 cm conditions; **Figure *6.4A right***). A bootstrap resampling procedure was performed to assess differences between groups, but no significant difference was found ($p < .01$). This result suggests that, contrary to

the original hypothesis, MSI does not vary as a function of perceived speaker size.

6.4.2.2 Unexpected PDR modulation by speaker size

Next, we examined whether speaker size information is reflected in the pupil dilation response (PDR). To test this, PDRs for the large and small size conditions were compared using bootstrap resampling. This analysis revealed a significant difference ($p < .01$) between the two conditions approximately 2.5–3 seconds after sound onset. Interestingly, the effect was in the opposite direction to the hypothesis: smaller-sized sounds elicited a larger PDR than larger-sized sounds (**Figure *6.4B***).

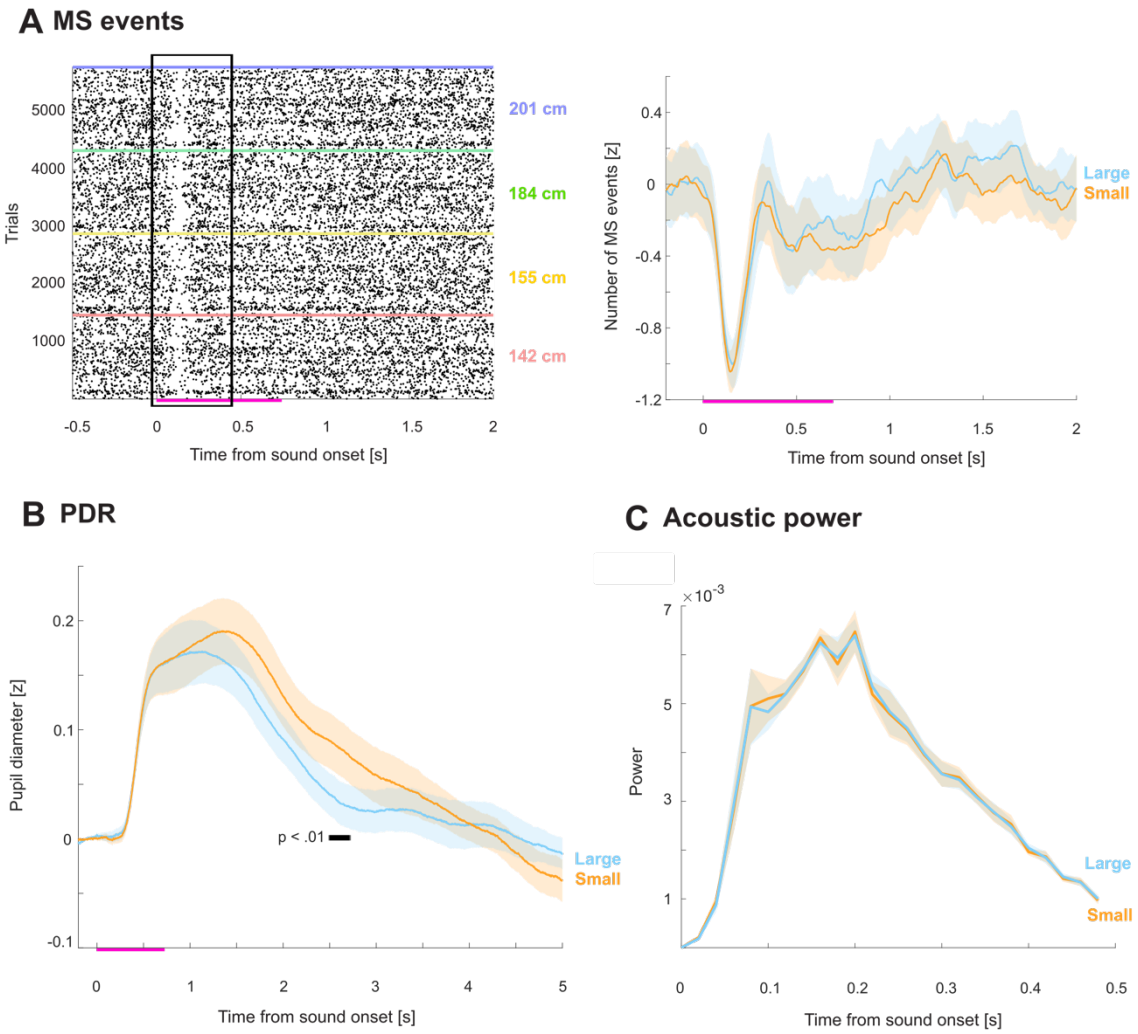


Figure 6.4 Ocular dynamics evoked by sounds conveying different speaker size.

[A] Microsaccadic (MS) events. Left: Raster plot of MS events pooled across all participants. Each dot represents the onset of an MS event, with the y-axis indicating individual trials grouped by speaker size conditions. The pink line along the x-axis marks the timing of sound presentation, and the black square highlights the microsaccadic inhibition (MSI). *Right:* Data on the left is smoothed and averaged for two size conditions (large: 201 cm and 184 cm conditions; small: 155 cm and 142 cm conditions). Shaded areas represent ± 2 standard errors of the mean (SEM). The pink line along the x-axis marks the timing of sound presentation. **[B]** Averaged pupil dilation response (PDR) for each size condition. Periods showing significant differences between conditions ($p < .01$, determined via bootstrap resampling) are indicated by bold horizontal lines. Shaded areas represent ± 2 SEM. The pink line along the x-axis marks the timing of sound presentation. **[C]** Average acoustic power for sounds representing small (142 cm and 155 cm) and large (184 cm and 201 cm) speaker sizes. Power was calculated in 20 ms bins over the first 500 ms of each stimulus. Shaded areas represent ± 2 SEM.

6.4.2.3 Acoustic power comparison across conditions

The stimuli used in this experiment were RMS-equalised to match their mean power. However, sounds with identical RMS values can still differ in perceived loudness due to variations in their long-term power profiles. Given that the PDR results contradicted the original hypothesis—and that the divergence appeared to emerge early in the PDR time course (though not significantly)—I tested whether differences in acoustic power might account for the observed effect.

To examine this, the first 500 ms of each sound was divided into 20 ms bins, and the average power within each bin was calculated (**Figure *6.4C***). Independent-samples t-tests ($\alpha = 0.05$) were conducted at each time bin to compare the large and small size conditions. No significant differences were found at any time point, even before correcting for multiple comparisons. These findings suggest that the two sound conditions were acoustically matched and that power differences are unlikely to explain the PDR result.

6.4.3 Discussion

The aim of this study was to investigate whether vocal size information is reflected in ocular dynamics, focusing specifically on microsaccadic inhibition (MSI) and pupil dilation response (PDR) as potential objective measures of auditory salience. I hypothesised that voices associated with larger speaker sizes would elicit stronger MSI and PDR responses. To test this, I conducted an online behavioural experiment and a controlled eye-tracking study. The online experiment confirmed participants' sensitivity to vocal size cues, showing reliable size discrimination across most conditions. However, contrary to the expectations, the eye-tracking results revealed no significant differences in MSI between sound conditions, and unexpectedly, smaller-sized voices evoked larger PDRs than larger-sized ones.

In the eye-tracking experiment, MSI was induced by all sound conditions (**Figure *6.4A***), yet no significant differences emerged between size conditions. I had hypothesised that larger-sized voices—presumably more salient due to their association with dominance or threat—would produce stronger MSI. The absence of this effect suggests that while the sound itself captures attention, differences in vocal size information do not modulate the degree of early attentional orienting as indexed by MSI.

Since microsaccades are primarily controlled by the superior colliculus (SC)—a midbrain structure involved in automatic attentional shifts (Hafed et al., 2009)—this null finding implies that size-related vocal cues may not be

represented at this subcortical level. This interpretation aligns with neuroimaging evidence indicating that speaker size perception involves higher-order cortical regions rather than subcortical structures (von Kriegstein et al., 2006, 2007). Thus, the lack of MSI modulation suggests that speaker size may not be an intrinsically salient feature processed through rapid, bottom-up mechanisms. Instead, it may rely on more abstract, interpretive processes requiring cortical integration and contextual evaluation. It is also possible that, in human voices, height does not necessarily convey threat or salience in the way it might in wild animals. For instance, Raine et al. (2018) found that judgments of strength and size of the speakers from the voices were uncorrelated, with listeners sensitive to both cues separately, implying that threat (strength)-related information may be embedded in other vocal features, not size. Alternatively, the size contrast between 142 cm and 201 cm may have been insufficient to elicit measurable differences in MSI responses. Future studies could leverage vocoder flexibility to manipulate vocal stimuli beyond typical human height ranges to more effectively test this hypothesis.

Alternatively, vocal size information captured bottom-up attention to some extent, but MSI was not sensitive enough to detect this. While MSI has been extensively studied in vision and shown to reflect visual salience (Bonneh et al., 2015; Rolfs et al., 2008), auditory evidence is limited—mostly stemming from one study linking MSI to roughness-mediated auditory salience (Zhao, Wai Yum, et al., 2019). Further research with a broader variety of auditory features is necessary to validate ocular dynamics as a reliable measure of objective sound salience.

Unlike MSI, the PDR showed a significant, though transient, difference between conditions: smaller-sized voices elicited larger pupil dilations, characterised by a more prolonged reduction phase following peak dilation compared to larger voices. This finding contradicts the hypothesis that larger voices would evoke stronger autonomic responses. Follow-up analyses confirmed that this effect was not explained by differences in acoustic power

during the early phase of stimuli. Although it is tempting to conclude that smaller-sized voices are more salient or arousing, it is also possible that perceived loudness—a subjective quality not fully controlled by RMS normalisation—contributed to the PDR pattern. Since loudness differences are known to affect pupil responses (Liao et al., 2016), future studies could incorporate explicit loudness ratings or use loudness roving to control for this factor.

Why did the PDR show a difference while MSI did not? This may reflect differences in their underlying neural circuits and temporal dynamics: MSI reflects early attentional orienting primarily mediated by the superior colliculus, whereas PDR indexes later, arousal-related processing linked to the locus coeruleus–norepinephrine system (Aston-Jones & Cohen, 2005; Contadini-Wright et al., 2023; Hafed et al., 2009; Joshi et al., 2016; Zhao, Wai Yum, et al., 2019). Alternatively, the discrepancy might be simply due to the inherently noisier nature of microsaccade data—since microsaccades occur only once or twice per second (Hafed et al., 2009), the necessary data smoothing could have obscured subtle condition differences that the more continuous pupil diameter measurements could capture.

Taken together, this study confirms that human listeners possess a remarkably robust ability to discriminate speaker size from vocal cues, even under minimal instruction and uncontrolled conditions. However, there was no clear modulation of ocular dynamics by speaker size, suggesting that size information may not serve as a bottom-up salience cue. The understanding of which auditory features capture bottom-up attention—and the use of ocular dynamics as objective measures of auditory salience—remains an emerging area of research. Further exploration of other factors contributing to auditory salience, alongside continued validation of ocular dynamics as reliable objective measure of auditory salience, will deepen our knowledge of auditory attention mechanisms and inform the design of improved soundscapes and auditory alarms.

References

Adams, R. P., & MacKay, D. J. C. (2007). *Bayesian Online Changepoint Detection* (No. arXiv:0710.3742). arXiv. https://doi.org/10.48550/arXiv.0710.3742

Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, *46*, 219–227. https://doi.org/10.1016/j.conb.2017.08.010

Amat, J., Baratta, M. V., Paul, E., Bland, S. T., Watkins, L. R., & Maier, S. F. (2005). Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. *Nature Neuroscience*, *8*(3), 365–371. https://doi.org/10.1038/nn1399

Andreou, L.-V., Kashino, M., & Chait, M. (2011). The role of temporal regularity in auditory segregation. *Hearing Research*, *280*(1), 228–235. https://doi.org/10.1016/j.heares.2011.06.001

Angel, L. A., Polzella, D. J., & Elvers, G. C. (2010). Background music and cognitive performance. *Perceptual and Motor Skills*, *110*(3 Pt 2), 1059–1064. https://doi.org/10.2466/pms.110.c.1059-1064

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A.-L., & Poeppel, D. (2015). Human Screams Occupy a Privileged Niche in the Communication Soundscape. *Current Biology*, *25*(15), 2051–2056. https://doi.org/10.1016/j.cub.2015.06.043

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Aslin, R. N., Jusczyk, P. W., & Pisoni, D. B. (1998). Speech and auditory processing during infancy: Constraints on and precursors to language. In *Handbook of child psychology: Volume 2: Cognition, perception, and language* (pp. 147–198). John Wiley & Sons, Inc.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological Psychiatry*, *46*(9), 1309–1320. https://doi.org/10.1016/s0006-3223(99)00140-7

Bach, D. R. (2014). Sympathetic nerve activity can be estimated from skin conductance responses—A comment on Henderson et al. (2012). *Neuroimage*, *84*, 122–123. https://doi.org/10.1016/j.neuroimage.2013.08.030

Bach, D. R., Daunizeau, J., Kuelzow, N., Friston, K. J., & Dolan, R. J. (2011). Dynamic causal modeling of spontaneous fluctuations in skin conductance. *Psychophysiology*, *48*(2), 252–257. https://doi.org/10.1111/j.1469-8986.2010.01052.x

Bach, D. R., Schächinger, H., Neuhoff, J. G., Esposito, F., Salle, F. D., Lehmann, C., Herdener, M., Scheffler, K., & Seifritz, E. (2008). Rising Sound Intensity: An Intrinsic Warning Cue Activating the Amygdala. *Cerebral Cortex*, *18*(1), 145–150. https://doi.org/10.1093/cercor/bhm040

Baijot, S., Slama, H., Söderlund, G., Dan, B., Deltenre, P., Colin, C., & Deconinck, N. (2016). Neuropsychological and neurophysiological benefits from white noise in children with and without ADHD. *Behavioral and Brain Functions : BBF*, *12*, 11. https://doi.org/10.1186/s12993-016-0095-y

Baldeweg, T. (2006). Repetition effects to sounds: Evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences*, *10*(3), 93–94. https://doi.org/10.1016/j.tics.2006.01.010

Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(5), E616–E625. https://doi.org/10.1073/pnas.1508523113

Basgol, H., Dayan, P., & Franz, V. H. (2025). Violation of auditory regularities is reflected in pupil dynamics. *Cortex*, *183*, 66–86. https://doi.org/10.1016/j.cortex.2024.10.023

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Bell, A. H., Summerfield, C., Morin, E. L., Malecek, N. J., & Ungerleider, L. G. (2016). Encoding of Stimulus Probability in Macaque Inferior Temporal Cortex. *Current Biology*, *26*(17), 2280–2290. https://doi.org/10.1016/j.cub.2016.07.007

Bendixen, A. (2014). Predictability effects in auditory scene analysis: A review. *Frontiers in Neuroscience*, *8*(8 MAR), 1–16. https://doi.org/10.3389/fnins.2014.00060

Bendixen, A., Roeber, U., & Schröger, E. (2007). Regularity extraction and application in dynamic auditory stimulus sequences. *Journal of Cognitive Neuroscience*, *19*(10), 1664–1677. https://doi.org/10.1162/jocn.2007.19.10.1664

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *83*(2), 120–131. https://doi.org/10.1016/j.ijpsycho.2011.08.003

Bendixen, A., & Schröger, E. (2008). Memory trace formation for abstract auditory features and its consequences in different attentional contexts. *Biological Psychology*, *78*(3), 231–241. https://doi.org/10.1016/j.biopsycho.2008.03.005

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience*, *29*(26), 8447–8451. https://doi.org/10.1523/JNEUROSCI.1493-09.2009

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Benjamin, L., Sablé-Meyer, M., Fló, A., Dehaene-Lambertz, G., & Al Roumi, F. (2024). Long-Horizon Associative Learning Explains Human Sensitivity to Statistical and Network Structures in Auditory Sequences. *The Journal of Neuroscience*, *44*(14), e1369232024. https://doi.org/10.1523/JNEUROSCI.1369-23.2024

Bianco, R., Harrison, P. M., Hu, M., Bolger, C., Picken, S., Pearce, M. T., & Chait, M. (2020). Long-term implicit memory for sequential auditory patterns in humans. *eLife*, *9*, e56073. https://doi.org/10.7554/eLife.56073

Bianco, R., Magami, K., Pearce, M., & Chait, M. (2025). *Discovery, Interruption, and Updating of Auditory Regularities in Memory: Evidence from Low-*

Frequency Brain Dynamics in Human MEG (p. 2025.03.28.645906). bioRxiv. https://doi.org/10.1101/2025.03.28.645906

Bidelman, G. M., & Myers, M. H. (2020). Frontal cortex selectively overrides auditory processing to bias perception for looming sonic motion. *Brain Research*, *1726*, 146507. https://doi.org/10.1016/j.brainres.2019.146507

Billig, A. J., Lad, M., Sedley, W., & Griffiths, T. D. (2022). The hearing hippocampus. *Progress in Neurobiology*, *218*, 102326. https://doi.org/10.1016/j.pneurobio.2022.102326

Bland, A. R., & Schaefer, A. (2012). Different varieties of uncertainty in human decision-making. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00085

Blank, H., & Davis, M. H. (2016). Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. *PLoS Biology*, *14*(11), e1002577. https://doi.org/10.1371/journal.pbio.1002577

Bonneh, Y. S., Adini, Y., & Polat, U. (2015). Contrast sensitivity revealed by microsaccades. *Journal of Vision*, *15*(9), 11. https://doi.org/10.1167/15.9.11

Bornstein, A. M., & Daw, N. D. (2012). Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience*, *35*(7), 1011–1023. https://doi.org/10.1111/j.1460-9568.2011.07920.x

Boubenec, Y., Lawlor, J., Górska, U., Shamma, S., & Englitz, B. (2017). Detecting changes in dynamic and complex acoustic environments. *eLife*, *6*, e24910. https://doi.org/10.7554/eLife.24910

Boucsein, W. (2012). *Electrodermal Activity*. Springer US. https://doi.org/10.1007/978-1-4614-1126-0

Bouret, S., & Sara, S. J. (2005). Network reset: A simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neurosciences*, *28*(11), 574–582. https://doi.org/10.1016/j.tins.2005.09.002

Bouwkamp, F. G., de Lange, F. P., & Spaak, E. (2025). Spatial Predictive Context Speeds Up Visual Search by Biasing Local Attentional Competition. *Journal of Cognitive Neuroscience*, *37*(1), 28–42. https://doi.org/10.1162/jocn_a_02254

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLOS ONE*, *5*(5), e10773. https://doi.org/10.1371/journal.pone.0010773

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596. https://doi.org/10.1038/nn.3961

Cheung, V. K. M., Harrison, P. M. C., Koelsch, S., Pearce, M. T., Friederici, A. D., & Meyer, L. (2023). Cognitive and sensory expectations independently shape musical expectancy and pleasure. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *379*(1895), 20220420. https://doi.org/10.1098/rstb.2022.0420

Cheung, V. K. M., Harrison, P. M. C., Meyer, L., Pearce, M. T., Haynes, J.-D., & Koelsch, S. (2019). Uncertainty and Surprise Jointly Predict Musical Pleasure and Amygdala, Hippocampus, and Auditory Cortex Activity. *Current Biology*, *29*(23), 4084-4092.e4. https://doi.org/10.1016/j.cub.2019.09.067

Clewett, D., DuBrow, S., & Davachi, L. (2019). Transcending time in the brain: How event memories are constructed from experience. *Hippocampus*, *29*(3), 162–183. https://doi.org/10.1002/hipo.23074

Clewett, D., Dunsmoor, J., Bachman, S. L., Phelps, E. A., & Davachi, L. (2022). Survival of the salient: Aversive learning rescues otherwise forgettable memories via neural reactivation and post-encoding hippocampal connectivity. *Neurobiology of Learning and Memory*, *187*, 107572. https://doi.org/10.1016/j.nlm.2021.107572

Clewett, D., Gasser, C., & Davachi, L. (2020). Pupil-linked arousal signals track the temporal organization of events in memory. *Nature Communications*, *11*(1), 4007. https://doi.org/10.1038/s41467-020-17851-9

Clewett, D., Huang, R., & Davachi, L. (2025). Locus coeruleus activation "resets" hippocampal event representations and separates adjacent memories. *Neuron*. https://doi.org/10.1016/j.neuron.2025.05.013

Clewett, D., & McClay, M. (2025). Emotional arousal lingers in time to bind discrete episodes in memory. *Cognition and Emotion*, *39*(1), 97–116. https://doi.org/10.1080/02699931.2023.2295853

Contadini-Wright, C., Magami, K., Mehta, N., & Chait, M. (2023). Pupil Dilation and Microsaccades Provide Complementary Insights into the Dynamics of Arousal and Instantaneous Attention during Effortful Listening. *Journal*

of Neuroscience, *43*(26), 4856–4866. https://doi.org/10.1523/JNEUROSCI.0242-23.2023

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Costa-Faidella, J., Baldeweg, T., Grimm, S., & Escera, C. (2011). Interactions between “What” and “When” in the Auditory System: Temporal Predictability Enhances Repetition Suppression. *The Journal of Neuroscience*, *31*(50), 18590–18597. https://doi.org/10.1523/JNEUROSCI.2599-11.2011

Coy, N., Bendixen, A., Grimm, S., Roeber, U., & Schröger, E. (2024). Conditional deviant repetition in the oddball paradigm modulates processing at the level of P3a but not MMN. *Psychophysiology*, *61*(6), e14545. https://doi.org/10.1111/psyp.14545

Critchley, H. D. (2002). Review: Electrodermal Responses: What Happens in the Brain. *The Neuroscientist*, *8*(2), 132–142. https://doi.org/10.1177/107385840200800209

Dahl, M. J., Mather, M., Sander, M. C., & Werkle-Bergner, M. (2020). Noradrenergic Responsiveness Supports Selective Attention across the Adult Lifespan. *Journal of Neuroscience*, *40*(22), 4372–4390. https://doi.org/10.1523/JNEUROSCI.0398-19.2020

Dahl, M. J., Mather, M., & Werkle-Bergner, M. (2022). Noradrenergic modulation of rhythmic neural activity shapes selective attention. *Trends in Cognitive Sciences*, *26*(1), 38–52. https://doi.org/10.1016/j.tics.2021.10.009

Dalton, P., & Lavie, N. (2004). Auditory attentional capture: Effects of singleton distractor sounds. *Journal of Experimental Psychology. Human Perception and Performance*, *30*(1), 180–193. https://doi.org/10.1037/0096-1523.30.1.180

Dawson, M. E., Schell, A. M., & Filion, D. L. (2016). The Electrodermal System. In G. G. Berntson, J. T. Cacioppo, & L. G. Tassinary (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 217–243). Cambridge University Press. https://doi.org/10.1017/9781107415782.010

de Berker, A. O., Rutledge, R. B., Mathys, C., Marshall, L., Cross, G. F., Dolan, R. J., & Bestmann, S. (2016). Computations of uncertainty mediate acute stress responses in humans. *Nature Communications*, *7*(1), 10996. https://doi.org/10.1038/ncomms10996

de Cheveigné, A., & Parra, L. C. (2014). Joint decorrelation, a versatile tool for multichannel data analysis. *NeuroImage*, *98*, 487–505. https://doi.org/10.1016/j.neuroimage.2014.05.068

de Cheveigné, A., & Simon, J. Z. (2008). Denoising based on spatial filtering. *Journal of Neuroscience Methods*, *171*(2), 331–339. https://doi.org/10.1016/j.jneumeth.2008.03.015

de Gee, J. W., Mridha, Z., Hudson, M., Shi, Y., Ramsaywak, H., Smith, S., Karediya, N., Thompson, M., Jaspe, K., Jiang, H., Zhang, W., & McGinley, M. J. (2024). Strategic stabilization of arousal boosts sustained attention. *Current Biology*, *34*(18), 4114-4128.e6. https://doi.org/10.1016/j.cub.2024.07.070

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Demarchi, G., Sanchez, G., & Weisz, N. (2019). Automatic and feature-specific prediction-related neural activity in the human auditory system. *Nature Communications*, *10*(1), 3440. https://doi.org/10.1038/s41467-019-11440-1

Devauges, V., & Sara, S. J. (1990). Activation of the noradrenergic system facilitates an attentional shift in the rat. *Behavioural Brain Research*, *39*(1), 19–28. https://doi.org/10.1016/0166-4328(90)90118-X

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

Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the late positive complex: Mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology*, *41*(5), 665–678. https://doi.org/10.1111/j.1469-8986.2004.00193.x

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology. General*, *142*(4), 1277–1286. https://doi.org/10.1037/a0034024

DuBrow, S., & Davachi, L. (2014). Temporal Memory Is Shaped by Encoding Stability and Intervening Item Reactivation. *The Journal of Neuroscience*, *34*(42), 13998–14005. https://doi.org/10.1523/JNEUROSCI.2535-14.2014

Dunn, M. A., Gomes, H., & Gravel, J. (2008). Mismatch negativity in children with autism and typical development. *Journal of Autism and Developmental Disorders*, *38*(1), 52–71. https://doi.org/10.1007/s10803-007-0359-3

Dunsmoor, J. E., Kroes, M. C. W., Moscatelli, C. M., Evans, M. D., Davachi, L., & Phelps, E. A. (2018). Event segmentation protects emotional memories from competing experiences encoded close in time. *Nature Human Behaviour*, *2*(4), 291–299. https://doi.org/10.1038/s41562-018-0317-4

Dunsmoor, J. E., Murty, V. P., Davachi, L., & Phelps, E. A. (2015). Emotional learning selectively and retroactively strengthens memories for related events. *Nature*, *520*(7547), 345–348. https://doi.org/10.1038/nature14106

Dürschmid, S., Edwards, E., Reichert, C., Dewar, C., Hinrichs, H., Heinze, H.-J., Kirsch, H. E., Dalal, S. S., Deouell, L. Y., & Knight, R. T. (2016). Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proceedings of the National Academy of Sciences*, *113*(24), 6755–6760. https://doi.org/10.1073/pnas.1525030113

Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429246593

Ekman, M., Kok, P., & de Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications*, *8*(1), 15276. https://doi.org/10.1038/ncomms15276

Ezzyat, Y., & Clements, A. (2024). Neural Activity Differentiates Novel and Learned Event Boundaries. *Journal of Neuroscience*, *44*(38). https://doi.org/10.1523/JNEUROSCI.2246-23.2024

Ezzyat, Y., & Davachi, L. (2011). What Constitutes an Episode in Episodic Memory? *Psychological Science*, *22*(2), 243–252. https://doi.org/10.1177/0956797610393742

Ezzyat, Y., & Davachi, L. (2014). Similarity Breeds Proximity: Pattern Similarity within and across Contexts Is Related to Later Mnemonic Judgments of Temporal Proximity. *Neuron*, *81*(5), 1179–1189. https://doi.org/10.1016/j.neuron.2014.01.042

Fairchild, L. (1981). Mate selection and behavioral thermoregulation in Fowler's toads. *Science*, *212*(4497), 950–951. https://doi.org/10.1126/science.212.4497.950

Fearnhead, P., & Liu, Z. (2007). On-Line Inference for Multiple Changepoint Problems. *Journal of the Royal Statistical Society Series B: Statistical*

Methodology, *69*(4), 589–605. https://doi.org/10.1111/j.1467-9868.2007.00601.x

Feldman, H., & Friston, K. (2010). Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00215

Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1-a

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*(3), 1511–1522. https://doi.org/10.1121/1.427148

Fitzgerald, K., & Todd, J. (2020). Making Sense of Mismatch Negativity. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/fpsyt.2020.00468

Flores, S., Bailey, H. R., Eisenberg, M. L., & Zacks, J. M. (2017). Event segmentation improves event memory up to one month later. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(8), 1183–1202. https://doi.org/10.1037/xlm0000367

Friederici, A. D., Mueller, J. L., & Oberecker, R. (2011). Precursors to Natural Grammar Learning: Preliminary Evidence from 4-Month-Old Infants. *PLOS ONE*, *6*(3), e17920. https://doi.org/10.1371/journal.pone.0017920

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2008). Hierarchical Models in the Brain. *PLOS Computational Biology*, *4*(11), e1000211. https://doi.org/10.1371/journal.pcbi.1000211

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Fritsche, M., Solomon, S. G., & de Lange, F. P. (2022). Brief Stimuli Cast a Persistent Long-Term Trace in Visual Cortex. *The Journal of Neuroscience*, *42*(10), 1999–2010. https://doi.org/10.1523/JNEUROSCI.1350-21.2021

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., & Kilner, J. M. (2008). The functional anatomy of the MMN: A DCM study of

the roving paradigm. *NeuroImage*, *42*(2), 936–944. https://doi.org/10.1016/j.neuroimage.2008.05.018

Garrido, M. I., Kilner, J. M., Kiebel, S. J., & Friston, K. J. (2007). Evoked brain responses are generated by feedback loops. *Proceedings of the National Academy of Sciences*, *104*(52), 20961–20966. https://doi.org/10.1073/pnas.0706274105

Garrido, M. I., Kilner, J. M., Kiebel, S. J., & Friston, K. J. (2009). Dynamic Causal Modeling of the Response to Frequency Deviants. *Journal of Neurophysiology*, *101*(5), 2620–2631. https://doi.org/10.1152/jn.90291.2008

Garrido, M. I., Kilner, J. M., Kiebel, S. J., Stephan, K. E., Baldeweg, T., & Friston, K. J. (2009). Repetition suppression and plasticity in the human brain. *NeuroImage*, *48*(1), 269–279. https://doi.org/10.1016/j.neuroimage.2009.06.034

Garrido, M. I., Kilner, J. M., Kiebel, S. J., Stephan, K. E., & Friston, K. J. (2007). Dynamic causal modelling of evoked potentials: A reproducibility study. *NeuroImage*, *36*(3), 571–580. https://doi.org/10.1016/j.neuroimage.2007.03.014

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Garrido, M. I., Sahani, M., & Dolan, R. J. (2013). Outlier Responses Reflect Sensitivity to Statistical Structure in the Human Brain. *PLoS Computational Biology*, *9*(3). https://doi.org/10.1371/journal.pcbi.1002999

Gelbard-Sagiv, H., Magidov, E., Sharon, H., Hendler, T., & Nir, Y. (2018). Noradrenaline Modulates Visual Perception and Late Visually Evoked Activity. *Current Biology*, *28*(14), 2239-2249.e6. https://doi.org/10.1016/j.cub.2018.05.051

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D., & Logothetis, N. K. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. *Current Biology*, *17*(5), 425–430. https://doi.org/10.1016/j.cub.2007.01.029

Gil-da-Costa, R., Braun, A., Lopes, M., Hauser, M. D., Carson, R. E., Herscovitch, P., & Martin, A. (2004). Toward an evolutionary perspective on conceptual representation: Species-specific calls activate visual and affective processing systems in the macaque. *Proceedings of the*

National Academy of Sciences, *101*(50), 17516–17521. https://doi.org/10.1073/pnas.0408077101

Glaze, C. M., Filipowicz, A. L. S., Kable, J. W., Balasubramanian, V., & Gold, J. I. (2018). A bias–variance trade-off governs individual differences in on-line learning in an unpredictable environment. *Nature Human Behaviour*, *2*(3), 213–224. https://doi.org/10.1038/s41562-018-0297-4

Glaze, C. M., Kable, J. W., & Gold, J. I. (2015). Normative evidence accumulation in unpredictable environments. *eLife*, *4*(AUGUST2015), 1–27. https://doi.org/10.7554/eLife.08825

Gold, D. A., Zacks, J. M., & Flores, S. (2017). Effects of cues to event segmentation on subsequent memory. *Cognitive Research: Principles and Implications*, *2*(1), 1. https://doi.org/10.1186/s41235-016-0043-2

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, *7*(5), 567–580. https://doi.org/10.1111/j.1467-7687.2004.00381.x

Gomot, M., Blanc, R., Clery, H., Roux, S., Barthelemy, C., & Bruneau, N. (2011). Candidate electrophysiological endophenotypes of hyper-reactivity to change in autism. *Journal of Autism and Developmental Disorders*, *41*(6), 705–714. https://doi.org/10.1007/s10803-010-1091-y

Grassi, M., & Darwin, C. J. (2006). The subjective duration of ramped and damped sounds. *Perception & Psychophysics*, *68*(8), 1382–1392. https://doi.org/10.3758/BF03193737

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, *94*, 149–165. https://doi.org/10.1016/j.jml.2016.11.001

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America*, *100*(2), 1111–1121. https://doi.org/10.1121/1.416296

Haake, A. B. (2011). Individual music listening in workplace settings: An exploratory survey of offices in the UK. *Musicae Scientiae*, *15*(1), 107–129. https://doi.org/10.1177/1029864911398065

Hafed, Z. M., Chen, C. Y., & Tian, X. (2015). Vision, perception, and attention through the lens of microsaccades: Mechanisms and implications. *Frontiers in Systems Neuroscience*, *9*. https://doi.org/10.3389/fnsys.2015.00167

Hafed, Z. M., & Clark, J. J. (2002). Microsaccades as an overt measure of covert attention shifts. *Vision Research*, *42*(22), 2533–2545. https://doi.org/10.1016/S0042-6989(02)00263-8

Hafed, Z. M., Goffart, L., & Krauzlis, R. J. (2009). A neural mechanism for microsaccade generation in the primate superior colliculus. *Science*, *323*(5916), 940–943. https://doi.org/10.1126/science.1166112

Harford, E. E., Holt, L. L., & Abel, T. J. (2024). Unveiling the development of human voice perception: Neurobiological mechanisms and pathophysiology. *Current Research in Neurobiology*, *6*, 100127. https://doi.org/10.1016/j.crneur.2024.100127

Harrison, P. M. C., Bianco, R., Chait, M., & Pearce, M. T. (2020). PPM-Decay: A computational model of auditory prediction with memory decay. *PLOS Computational Biology*, *16*(11), e1008304. https://doi.org/10.1371/journal.pcbi.1008304

Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, *389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

Herrmann, B., Araz, K., & Johnsrude, I. S. (2021). Sustained neural activity correlates with rapid perceptual learning of auditory patterns. *NeuroImage*, *238*, 118238. https://doi.org/10.1016/j.neuroimage.2021.118238

Herrmann, B., Buckland, C., & Johnsrude, I. S. (2019). Neural signatures of temporal regularity processing in sounds differ between younger and older adults. *Neurobiology of Aging*, *83*, 73–85. https://doi.org/10.1016/j.neurobiolaging.2019.08.028

Herrmann, B., Henry, M. J., Fromboluti, E. K., McAuley, J. D., & Obleser, J. (2015). Statistical context shapes stimulus-specific adaptation in human auditory cortex. *Journal of Neurophysiology*, *113*(7), 2582–2591. https://doi.org/10.1152/jn.00634.2014

Herrmann, B., & Johnsrude, I. S. (2018). Neural Signatures of the Processing of Temporal Patterns in Sound. *Journal of Neuroscience*, *38*(24), 5466–5477. https://doi.org/10.1523/JNEUROSCI.0346-18.2018

Heusser, A. C., Ezzyat, Y., Shiff, I., & Davachi, L. (2018). Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *44*(7), 1075–1090. https://doi.org/10.1037/xlm0000503

Hodson, A. J., Shinn-Cunningham, B. G., & Holt, L. L. (2023). Statistical learning across passive listening adjusts perceptual weights of speech input dimensions. *Cognition*, *238*, 105473. https://doi.org/10.1016/j.cognition.2023.105473

Holt, L. L. (2025). Speech Perception Is Speech Learning. *Current Directions in Psychological Science*, 09637214251318726. https://doi.org/10.1177/09637214251318726

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisitiona). *The Journal of the Acoustical Society of America*, *119*(5), 3059–3071. https://doi.org/10.1121/1.2188377

Holt, L. L., Tierney, A. T., Guerra, G., Laffere, A., & Dick, F. (2018). Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. *Hearing Research*, *366*, 50–64. https://doi.org/10.1016/j.heares.2018.06.014

Horner, A. J., Bisby, J. A., Wang, A., Bogus, K., & Burgess, N. (2016). The role of spatial boundaries in shaping long-term event representations. *Cognition*, *154*, 151–164. https://doi.org/10.1016/j.cognition.2016.05.013

Horváth, J., Czigler, I., Sussman, E., & Winkler, I. (2001). Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. *Cognitive Brain Research*, *12*(1), 131–144. https://doi.org/10.1016/S0926-6410(01)00038-6

Hu, M., Bianco, R., Hidalgo, A. R., & Chait, M. (2024). Concurrent Encoding of Sequence Predictability and Event-Evoked Prediction Error in Unfolding Auditory Patterns. *The Journal of Neuroscience*, *44*(14), e1894232024. https://doi.org/10.1523/JNEUROSCI.1894-23.2024

Huang, N., & Elhilali, M. (2017). Auditory salience using natural soundscapes. *The Journal of the Acoustical Society of America*, *141*(3), 2163–2176. https://doi.org/10.1121/1.4979055

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *WIREs Cognitive Science*, *2*(5), 580–593. https://doi.org/10.1002/wcs.142

Idemaru, K., & Holt, L. L. (2011). Word Recognition Reflects Dimension-based Statistical Learning. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(6), 1939–1956. https://doi.org/10.1037/a0025641

Idemaru, K., & Holt, L. L. (2013). The developmental trajectory of children's perception and production of English /r/-/l/. *The Journal of the Acoustical*

Society of America, *133*(6), 4232–4246. https://doi.org/10.1121/1.4802905

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57. https://doi.org/10.1016/S0010-0277(02)00198-1

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

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation trade-off: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, *23*(7), 1587–1596. https://doi.org/10.1162/jocn.2010.21548

Johnston, R., Snyder, A. C., Khanna, S. B., Issar, D., & Smith, M. A. (2022). The eyes reflect an internal cognitive state hidden in the population activity of cortical neurons. *Cerebral Cortex (New York, N.Y.: 1991)*, *32*(15), 3331–3346. https://doi.org/10.1093/cercor/bhab418

Joshi, S., & Gold, J. I. (2020). Pupil Size as a Window on Neural Substrates of Cognition. *Trends in Cognitive Sciences*, *24*(6), 466–480. https://doi.org/10.1016/j.tics.2020.03.005

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*(1), 221–234. https://doi.org/10.1016/j.neuron.2015.11.028

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 675–687.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants′ Sensitivity to the Sound Patterns of Native Language Words. *Journal of Memory and Language*, *32*(3), 402–420. https://doi.org/10.1006/jmla.1993.1022

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants′ Sensitivity to Phonotactic Patterns in the Native Language. *Journal of Memory and Language*, *33*(5), 630–645. https://doi.org/10.1006/jmla.1994.1030

Kaczkurkin, A. N., Burton, P. C., Chazin, S. M., Manbeck, A. B., Espensen-Sturges, T., Cooper, S. E., Sponheim, S. R., & Lissek, S. (2017). Neural Substrates of Overgeneralized Conditioned Fear in PTSD. *The American Journal of Psychiatry*, *174*(2), 125–134. https://doi.org/10.1176/appi.ajp.2016.15121549

Kalbe, F., & Schwabe, L. (2020). Beyond arousal: Prediction error related to aversive events promotes episodic memory formation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(2), 234–246. https://doi.org/10.1037/xlm0000728

Kao, C.-H., Khambhati, A. N., Bassett, D. S., Nassar, M. R., McGuire, J. T., Gold, J. I., & Kable, J. W. (2020). Functional brain network reconfiguration during learning in a dynamic environment. *Nature Communications*, *11*(1), 1682. https://doi.org/10.1038/s41467-020-15442-2

Kawahara, H., & Irino, T. (2004). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In *Speech Separation by Humans and Machines* (pp. 167–180). Kluwer Academic.

Kawahara, H., Masuda-Katsuse, I., & De Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication*, *27*(3), 187–207. https://doi.org/10.1016/S0167-6393(98)00085-5

Kaya, E. M., & Elhilali, M. (2014). Investigating bottom-up auditory attention. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00327

Kaya, E. M., Huang, N., & Elhilali, M. (2020). Pitch, Timbre and Intensity Interdependently Modulate Neural Responses to Salient Sounds. *Neuroscience*, *440*, 1–14. https://doi.org/10.1016/j.neuroscience.2020.05.018

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2006). Memory for specific visual details can be enhanced by negative arousing content. *Journal of Memory and Language*, *54*(1), 99–112. https://doi.org/10.1016/j.jml.2005.05.005

Kern, P., Heilbron, M., de Lange, F. P., & Spaak, E. (2022). Cortical activity during naturalistic music listening reflects short-range predictions based on long-term experience. *eLife*, *11*, e80935. https://doi.org/10.7554/eLife.80935

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology*, *55*(Volume 55, 2004), 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Khouri, L., & Nelken, I. (2015). Detecting the unexpected. *Current Opinion in Neurobiology*, *35*, 142–147. https://doi.org/10.1016/j.conb.2015.08.003

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A Hierarchy of Time-Scales and the Brain. *PLOS Computational Biology*, *4*(11), e1000209. https://doi.org/10.1371/journal.pcbi.1000209

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Kohl, P. K. (1993). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics*, *21*(1), 125–139. https://doi.org/10.1016/S0095-4470(19)31326-9

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, *26*(7), 1546–1554. https://doi.org/10.1162/jocn_a_00562

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences*, *114*(39), 10473–10478. https://doi.org/10.1073/pnas.1705652114

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex (New York, N.Y.: 1991)*, *22*(9), 2197–2206. https://doi.org/10.1093/cercor/bhr310

Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, *124*(6), 3959–3971. https://doi.org/10.1121/1.2999341

Koolhaas, J. M., Bartolomucci, A., Buwalda, B., de Boer, S. F., Flügge, G., Korte, S. M., Meerlo, P., Murison, R., Olivier, B., Palanza, P., Richter-Levin, G., Sgoifo, A., Steimer, T., Stiedl, O., van Dijk, G., Wöhr, M., & Fuchs, E. (2011). Stress revisited: A critical evaluation of the stress concept. *Neuroscience & Biobehavioral Reviews*, *35*(5), 1291–1301. https://doi.org/10.1016/j.neubiorev.2011.02.003

Korn, C. W., Staib, M., Tzovara, A., Castegnetti, G., & Bach, D. R. (2017). A pupil size response model to assess fear learning. *Psychophysiology*, *54*(3), 330–343. https://doi.org/10.1111/psyp.12801

Krauzlis, R. J., Bogadhi, A. R., Herman, J. P., & Bollimunta, A. (2019). *Selective attention without a neocortex*. 161–175. https://doi.org/10.1016/j.cortex.2017.08.026.Selective

Krauzlis, R. J., Lovejoy, L. P., & Zénon, A. (2013). Superior colliculus and visual spatial attention. *Annual Review of Neuroscience*, *36*, 165–182. https://doi.org/10.1146/annurev-neuro-062012-170249

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science*, *255*(5044), 606–608. https://doi.org/10.1126/science.1736364

Kumar, S., Kaposvari, P., & Vogels, R. (2017). Encoding of Predictable and Unpredictable Stimuli by Inferior Temporal Cortical Neurons. *Journal of Cognitive Neuroscience*, *29*(8), 1445–1454. https://doi.org/10.1162/jocn_a_01135

Laretzaki, G., Plainis, S., Argyropoulos, S., Pallikaris, I. G., & Bitsios, P. (2010). Threat and anxiety affect visual contrast perception. *Journal of Psychopharmacology*, *24*(5), 667–675. https://doi.org/10.1177/0269881108098823

Lawson, R. P., Bisby, J., Nord, C. L., Burgess, N., & Rees, G. (2021). The Computational, Pharmacological, and Physiological Determinants of Sensory Learning under Uncertainty. *Current Biology*, *31*(1), 163-172.e4. https://doi.org/10.1016/j.cub.2020.10.043

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*, *20*(9), 1293–1299. https://doi.org/10.1038/nn.4615

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00302

Lecaignard, F., Bertrand, O., Caclin, A., & Mattout, J. (2022). Neurocomputational Underpinnings of Expected Surprise. *Journal of Neuroscience*, *42*(3), 474–486. https://doi.org/10.1523/JNEUROSCI.0601-21.2021

Lee, T. H., Greening, S. G., Ueno, T., Clewett, D., Ponzio, A., Sakaki, M., & Mather, M. (2018). Arousal increases neural gain via the locus coeruleus–noradrenaline system in younger adults but not in older adults. *Nature Human Behaviour*, *2*(5), 356–366. https://doi.org/10.1038/s41562-018-0344-1

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *JOSA A*, *20*(7), 1434–1448. https://doi.org/10.1364/JOSAA.20.001434

Leuchs, L., Schneider, M., & Spoormaker, V. I. (2019). Measuring the conditioned response: A comparison of pupillometry, skin conductance, and startle electromyography. *Psychophysiology*, *56*(1), e13283. https://doi.org/10.1111/psyp.13283

Liao, H. I., Kidani, S., Yoneya, M., Kashino, M., & Furukawa, S. (2016). Correspondences among pupillary dilation response, subjective salience of sounds, and loudness. *Psychonomic Bulletin and Review*, *23*(2), 412–425. https://doi.org/10.3758/s13423-015-0898-0

Lieder, I., Adam, V., Frenkel, O., Jaffe-Dax, S., Sahani, M., & Ahissar, M. (2019). Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nature Neuroscience*, *22*(2), 256–264. https://doi.org/10.1038/s41593-018-0308-9

Magami, K., Bianco, R., Hall, E., Pearce, M., & Chait, M. (2025). *The Effect of Previously Encountered Sensory Information on Neural Representations of Predictability: Evidence from Human EEG* (p. 2025.05.27.656332). bioRxiv. https://doi.org/10.1101/2025.05.27.656332

Maheu, M., Dehaene, S., & Meyniel, F. (2019). Brain signatures of a multiscale process of sequence learning in humans. *eLife*, *8*, 1–24. https://doi.org/10.7554/eLife.41541

Manning, J. R., & Kahana, M. J. (2012). Interpreting semantic clustering effects in free recall. *Memory*, *20*(5), 511–517. https://doi.org/10.1080/09658211.2012.683010

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule Learning by Seven-Month-Old Infants. *Science*, *283*(5398), 77–80. https://doi.org/10.1126/science.283.5398.77

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1), 16. https://doi.org/10.5334/joc.18

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111. https://doi.org/10.1016/S0010-0277(01)00157-3

McClay, M., Sachs, M. E., & Clewett, D. (2023). Dynamic emotional states shape the episodic structure of memory. *Nature Communications*, *14*(1), 6533. https://doi.org/10.1038/s41467-023-42241-2

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron*, *84*(4), 870–881. https://doi.org/10.1016/j.neuron.2014.10.013

Miller, G. A. (1947). Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *Journal of the Acoustical Society of America*, *19*, 609–619. https://doi.org/10.1121/1.1916528

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, *53*(4), 1551–1562. https://doi.org/10.3758/s13428-020-01514-0

Milne, A. E., Zhao, S., Tampakaki, C., Bury, G., & Chait, M. (2021). Sustained pupil responses are modulated by predictability of auditory sequences. *Journal of Neuroscience*, *41*(28), 6116–6127. https://doi.org/10.1523/JNEUROSCI.2879-20.2021

Morise, M., Yokomori, F., & Ozawa, K. (2016). WORLD : A vocoder-based high-quality speech synthesis system for real-time applications. *IEICE TRANS.INF.&SYST*, *99*, 1877–1884.

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, *66*(3), 241–251. https://doi.org/10.1007/BF00198477

Murphy, S., Fraenkel, N., & Dalton, P. (2013). Perceptual load does not modulate auditory distractor processing. *Cognition*, *129*(2), 345–355. https://doi.org/10.1016/j.cognition.2013.07.014

Murphy, T. K., Nozari, N., & Holt, L. (2024). Transfer of statistical learning from passive speech perception to speech production. *Psychonomic Bulletin & Review*, *31*(3), 1193–1205. https://doi.org/10.3758/s13423-023-02399-8

Murty, V. P., & Adcock, R. A. (2014). Enriched encoding: Reward motivation organizes cortical networks for hippocampal detection of unexpected events. *Cerebral Cortex (New York, N.Y.: 1991)*, *24*(8), 2160–2168. https://doi.org/10.1093/cercor/bht063

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*(1), 1–21. https://doi.org/10.1111/1469-8986.3810001

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*(6), 826–859. https://doi.org/10.1037/0033-2909.125.6.826

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. https://doi.org/10.1038/nn.3130

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, *30*(37), 12366–12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010

Natan, R. G., Briguglio, J. J., Mwilambwe-Tshilobo, L., Jones, S. I., Aizenberg, M., Goldberg, E. M., & Geffen, M. N. (2015). Complementary control of sensory adaptation by two types of cortical interneurons. *eLife*, *4*, e09868. https://doi.org/10.7554/eLife.09868

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology*, *13*(2), 87–110. https://doi.org/10.1207/S15326969ECO1302_2

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2

Ngo, M. K., & Spence, C. (2010). Crossmodal facilitation of masked visual target identification. *Attention, Perception, & Psychophysics*, *72*(7), 1938–1947. https://doi.org/10.3758/APP.72.7.1938

Nobre, A., Correa, A., & Coull, J. (2007). The hazards of time. *Current Opinion in Neurobiology*, *17*(4), 465–470. https://doi.org/10.1016/j.conb.2007.07.006

Norman-Haignere, S. V., Long, L. K., Devinsky, O., Doyle, W., Irobunda, I., Merricks, E. M., Feldstein, N. A., McKhann, G. M., Schevon, C. A., Flinker, A., & Mesgarani, N. (2022). Multiscale temporal integration organizes hierarchical computation in human auditory cortex. *Nature Human Behaviour*, *6*(3), 455–469. https://doi.org/10.1038/s41562-021-01261-y

Nutt, D. J., & Malizia, A. L. (2004). Structural and functional brain changes in posttraumatic stress disorder. *The Journal of Clinical Psychiatry*, *65 Suppl 1*, 11–17.

Ojala, K. E., & Bach, D. R. (2020). Measuring learning in human classical threat conditioning: Translational, cognitive and methodological considerations.

Neuroscience & Biobehavioral Reviews, *114*, 96–112. https://doi.org/10.1016/j.neubiorev.2020.04.019

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D. Y., & Schröger, E. (2002). Differential Contribution of Frontal and Temporal Cortices to Auditory Change Detection: fMRI and ERP Results. *NeuroImage*, *15*(1), 167–174. https://doi.org/10.1006/nimg.2001.0970

O'Reilly, J. X. (2013). Making predictions in a changing world-inference, uncertainty, and learning. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00105

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology*, *88*(2), 109–123. https://doi.org/10.1016/j.ijpsycho.2013.03.015

Paavilainen, P., Arajärvi, P., & Takegata, R. (2007). Preattentive detection of nonsalient contingencies between auditory features. *NeuroReport*, *18*(2), 159–163. https://doi.org/10.1097/WNR.0b013e328010e2ac

Padmala, S., & Pessoa, L. (2008). Affective Learning Enhances Visual Detection and Responses in Primary Visual Cortex. *Journal of Neuroscience*, *28*(24), 6202–6210. https://doi.org/10.1523/JNEUROSCI.1233-08.2008

Paraskevopoulos, E., Chalas, N., Kartsidis, P., Wollbrink, A., & Bamidis, P. (2018). Statistical learning of multisensory regularities is enhanced in musicians: An MEG study. *NeuroImage*, *175*(January), 150–160. https://doi.org/10.1016/j.neuroimage.2018.04.002

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107–123. https://doi.org/10.1016/S0042-6989(01)00250-4

Payzan-LeNestour, E., & Bossaerts, P. (2011). Risk, Unexpected Uncertainty, and Estimation Uncertainty: Bayesian Learning in Unstable Settings. *PLOS Computational Biology*, *7*(1), e1001048. https://doi.org/10.1371/journal.pcbi.1001048

Payzan-LeNestour, E., Dunne, S., Bossaerts, P., & O'Doherty, J. P. (2013). The neural representation of unexpected uncertainty during value-based decision making. *Neuron*, *79*(1), 191–201. https://doi.org/10.1016/j.neuron.2013.04.037

Pearce, M. T. (2005). The construction and evaluation of statistical models of melodic structure in music perception and composition. *PhD Thesis, City, University of London*. http://openaccess.city.ac.uk/1189/

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, *1423*(1), 378–395. https://doi.org/10.1111/nyas.13654

Peel, T. R., Hafed, Z. M., Dash, S., Lomber, S. G., & Corneil, B. D. (2016). A Causal Role for the Cortical Frontal Eye Fields in Microsaccade Deployment. *PLoS Biology*, *14*(8), 1–23. https://doi.org/10.1371/journal.pbio.1002531

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development*, *80*(3), 10.1111/j.1467-8624.2009.01290.x. https://doi.org/10.1111/j.1467-8624.2009.01290.x

Peters, A., McEwen, B. S., & Friston, K. (2017). Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. *Progress in Neurobiology*, *156*, 164–188. https://doi.org/10.1016/j.pneurobio.2017.05.004

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*(18), 2397–2416. https://doi.org/10.1016/j.visres.2005.03.019

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience*, *11*(3), 367–374. https://doi.org/10.1038/nn2043

Pettijohn, K. A., & Radvansky, G. A. (2016). Narrative event boundaries, reading times, and expectation. *Memory & Cognition*, *44*(7), 1064–1075. https://doi.org/10.3758/s13421-016-0619-6

Pettijohn, K. A., Thompson, A. N., Tamplin, A. K., Krawietz, S. A., & Radvansky, G. A. (2016). Event boundaries and memory improvement. *Cognition*, *148*, 136–144. https://doi.org/10.1016/j.cognition.2015.12.013

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion Facilitates Perception and Potentiates the Perceptual Benefits of Attention. *Psychological Science*, *17*(4), 292. https://doi.org/10.1111/j.1467-9280.2006.01701.x

Phillips, H. N., Blenkmann, A., Hughes, L. E., Bekinschtein, T. A., & Rowe, J. B. (2015). Hierarchical Organization of Frontotemporal Networks for the Prediction of Stimuli across Multiple Dimensions. *The Journal of*

Neuroscience: The Official Journal of the Society for Neuroscience, *35*(25), 9255–9264. https://doi.org/10.1523/JNEUROSCI.5095-14.2015

Phillips, H. N., Blenkmann, A., Hughes, L. E., Kochen, S., Bekinschtein, T. A., Cam-Can, null, & Rowe, J. B. (2016). Convergent evidence for hierarchical prediction networks from human electrocorticography and magnetoencephalography. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *82*, 192–205. https://doi.org/10.1016/j.cortex.2016.05.001

Piray, P., & Daw, N. D. (2024). Computational processes of simultaneous learning of stochasticity and volatility in humans. *Nature Communications*, *15*(1), 9073. https://doi.org/10.1038/s41467-024-53459-z

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., Röder, S., Andrews, P. W., Fink, B., DeBruine, L. M., Jones, B. C., & Feinberg, D. R. (2014). Vocal indicators of body size in men and women: A meta-analysis. *Animal Behaviour*, *95*, 89–99. https://doi.org/10.1016/j.anbehav.2014.06.011

Ponsot, E., Susini, P., & Meunier, S. (2015). A robust asymmetry in loudness between rising- and falling-intensity tones. *Attention, Perception, & Psychophysics*, *77*(3), 907–920. https://doi.org/10.3758/s13414-014-0824-y

Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science*, *357*(6351), 596–600. https://doi.org/10.1126/science.aan3458

Press, C., Kok, P., & Yon, D. (2020). The Perceptual Prediction Paradox. *Trends in Cognitive Sciences*, *24*(1), 13–24. https://doi.org/10.1016/j.tics.2019.11.003

Pu, Y., Kong, X.-Z., Ranganath, C., & Melloni, L. (2022). Event boundaries shape temporal organization of memory by resetting temporal context. *Nature Communications*, *13*, 622. https://doi.org/10.1038/s41467-022-28216-9

Quiroga-Martinez, D. R., Hansen, N. C., Højlund, A., Pearce, M., Brattico, E., Holmes, E., Friston, K., & Vuust, P. (2021). Musicianship and melodic predictability enhance neural gain in auditory cortex during pitch deviance detection. *Human Brain Mapping*, *42*(17), 5595–5608. https://doi.org/10.1002/hbm.25638

Raccah, O., Doelling, K. B., Davachi, L., & Poeppel, D. (2023). Acoustic features drive event segmentation in speech. *Journal of Experimental*

Psychology. Learning, Memory, and Cognition, *49*(9), 1494–1504. https://doi.org/10.1037/xlm0001150

Radvansky, G. A., Tamplin, A. K., Armendarez, J., & Thompson, A. N. (2014). Different Kinds of Causality in Event Cognition. *Discourse Processes: A Multidisciplinary Journal*, *51*(7), 601–618. https://doi.org/10.1080/0163853X.2014.903366

Raine, J., Pisanski, K., Oleszkiewicz, A., Simner, J., & Reby, D. (2018). Human Listeners Can Accurately Judge Strength and Height Relative to Self from Aggressive Roars and Speech. *iScience*, *4*, 273–280. https://doi.org/10.1016/j.isci.2018.05.002

Raio, C. M., Hartley, C. A., Orederu, T. A., Li, J., & Phelps, E. A. (2017). Stress attenuates the flexible updating of aversive value. *Proceedings of the National Academy of Sciences*, *114*(42), 11241–11246. https://doi.org/10.1073/pnas.1702565114

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1566), 941–947. https://doi.org/10.1098/rspb.2004.2954

Reisinger, L., Demarchi, G., Obleser, J., Sedley, W., Partyka, M., Schubert, J., Gehmacher, Q., Roesch, S., Suess, N., Trinka, E., Schlee, W., & Weisz, N. (2024). Aberrant auditory prediction patterns robustly characterize tinnitus. *eLife*, *13*, RP99757. https://doi.org/10.7554/eLife.99757

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A Computational Model of Event Segmentation From Perceptual Prediction. *Cognitive Science*, *31*(4), 613–643. https://doi.org/10.1080/15326900701399913

Rolfs, M. (2009). Microsaccades: Small steps on a long way. *Vision Research*, *49*(20), 2415–2441. https://doi.org/10.1016/j.visres.2009.08.010

Rolfs, M., Kliegl, R., & Engbert, R. (2008). Toward a model of microsaccade generation: The case of microsaccadic inhibition. *Journal of Vision*, *8*(11), 1–23. https://doi.org/10.1167/8.11.5

Rouhani, N., Norman, K. A., & Niv, Y. (2018). Dissociable effects of surprising rewards on learning and memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *44*(9), 1430–1443. https://doi.org/10.1037/xlm0000518

Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2020). Reward prediction errors create event boundaries in memory. *Cognition*, *203*, 104269. https://doi.org/10.1016/j.cognition.2020.104269

Rubin, J., Ulanovsky, N., Nelken, I., & Tishby, N. (2016). The Representation of Prediction Error in Auditory Cortex. *PLOS Computational Biology*, *12*(8), e1005058. https://doi.org/10.1371/journal.pcbi.1005058

Rucci, M., & Poletti, M. (2015). Control and functions of fixational eye movements. *Annual Review of Vision Science*, *1*, 499–518. https://doi.org/10.1146/annurev-vision-082114-035742

Saberi, K., & Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature*, *398*.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*, *14*(2), 257–262. https://doi.org/10.1038/nn.2726

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, *10*(3), 211–223. https://doi.org/10.1038/nrn2573

Sara, S. J., & Bouret, S. (2012). Orienting and reorienting: The locus coeruleus mediates cognition through arousal. *Neuron*, *76*(1), 130–141. https://doi.org/10.1016/j.neuron.2012.09.011

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., Eisenberg, M. L., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241–255. https://doi.org/10.1016/j.cognition.2013.07.002

Sartory, G., Cwik, J., Knuppertz, H., Schürholt, B., Lebens, M., Seitz, R. J., & Schulze, R. (2013). In search of the trauma memory: A meta-analysis of functional neuroimaging studies of symptom provocation in posttraumatic

stress disorder (PTSD). *PloS One*, *8*(3), e58150. https://doi.org/10.1371/journal.pone.0058150

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The Necessity of the Medial Temporal Lobe for Statistical Learning. *Journal of Cognitive Neuroscience*, *26*(8), 1736–1747. https://doi.org/10.1162/jocn_a_00578

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities. *Current Biology*, *22*(17), 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486–492. https://doi.org/10.1038/nn.3331

Schröger, E., & Roeber, U. (2021). Encoding of deterministic and stochastic auditory rules in the human brain: The mismatch negativity mechanism does not reflect basic probability. *Hearing Research*, *399*, 107907. https://doi.org/10.1016/j.heares.2020.107907

Schubert, J., Schmidt, F., Gehmacher, Q., Bresgen, A., & Weisz, N. (2023). Cortical speech tracking is related to individual prediction tendencies. *Cerebral Cortex*, 1–12. https://doi.org/10.1093/cercor/bhac528

Schulz, A., Miehl, C., Berry, M. J., II, & Gjorgjieva, J. (2021). The generation of cortical novelty responses through inhibitory plasticity. *eLife*, *10*, e65309. https://doi.org/10.7554/eLife.65309

Sedley, W., Friston, K. J., Gander, P. E., Kumar, S., & Griffiths, T. D. (2016). An integrative tinnitus model based on sensory precision. *Trends in Neurosciences*, *39*(12), 799–812. https://doi.org/10.1016/j.tins.2016.10.004

Sherman, B. E., DuBrow, S., Winawer, J., & Davachi, L. (2023). Mnemonic Content and Hippocampal Patterns Shape Judgments of Time. *Psychological Science*, *34*(2), 221–237. https://doi.org/10.1177/09567976221129533

Shipp, S. (2016). Neural elements for predictive coding. *Frontiers in Psychology*, *7*(NOV), 1–21. https://doi.org/10.3389/fpsyg.2016.01792

Siefke, B. M., Smith, T. A., & Sederberg, P. B. (2019). A context-change account of temporal distinctiveness. *Memory & Cognition*, *47*(6), 1158–1172. https://doi.org/10.3758/s13421-019-00925-5

Sirois, S., & Brisson, J. (2014). Pupillometry. *WIREs Cognitive Science*, *5*(6), 679–692. https://doi.org/10.1002/wcs.1323

Skerritt-Davis, B., & Elhilali, M. (2018). Detecting change in stochastic sound sequences. *PLOS Computational Biology*, *14*(5), e1006162. https://doi.org/10.1371/journal.pcbi.1006162

Skerritt-Davis, B., & Elhilali, M. (2021a). Computational framework for investigating predictive processing in auditory perception. *Journal of Neuroscience Methods*, *360*, 109177. https://doi.org/10.1016/j.jneumeth.2021.109177

Skerritt-Davis, B., & Elhilali, M. (2021b). Neural encoding of auditory statistics. *Journal of Neuroscience*, *41*(31), 6726–6739. https://doi.org/10.1523/JNEUROSCI.1887-20.2021

Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, *118*(5), 3177–3186. https://doi.org/10.1121/1.2047107

Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *J Acoust Soc Am*, *117*(1), 305–318.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Sohoglu, E., & Chait, M. (2016). Detecting and representing predictable structure during auditory scene analysis. *eLife*, *5*, 1–17. https://doi.org/10.7554/eLife.19113

Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. (2017). Event Boundaries Trigger Rapid Memory Reinstatement of the Prior Events to Promote Their Representation in Long-Term Memory. *Current Biology*, *27*(22), 3499-3504.e4. https://doi.org/10.1016/j.cub.2017.09.057

Southwell, R., Baumann, A., Gal, C., Barascud, N., Friston, K., & Chait, M. (2017). Is predictability salient? A study of attentional capture by auditory patterns. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714). https://doi.org/10.1098/rstb.2016.0105

Southwell, R., & Chait, M. (2018). Enhanced deviant responses in patterned relative to random sound sequences. *Cortex*, *109*, 92–103. https://doi.org/10.1016/j.cortex.2018.08.032

Stefanics, G., Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: A predictive coding view. *Frontiers in Human Neuroscience*, *8*(September), 1–19. https://doi.org/10.3389/fnhum.2014.00666

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, *8*(6), 497–506. https://doi.org/10.1162/jocn.1996.8.6.497

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT.

Swallow, K. M., Jiang, Y. V., & Riley, E. B. (2019). Target detection increases pupil diameter and enhances memory for background scenes during multi-tasking. *Scientific Reports*, *9*(1), 5255. https://doi.org/10.1038/s41598-019-41658-4

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event Boundaries in Perception Affect Memory Encoding and Updating. *Journal of Experimental Psychology. General*, *138*(2), 236. https://doi.org/10.1037/a0015631

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*(1), 21. https://doi.org/10.1186/1471-2202-10-21

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. https://doi.org/10.1037/0012-1649.39.4.706

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America*, *85*(4), 1699–1707. https://doi.org/10.1121/1.397959

Tivadar, R. I., Knight, R. T., & Tzovara, A. (2021). Automatic Sensory Predictions: A Review of Predictive Mechanisms in the Brain and Their Link to Conscious Processing. *Frontiers in Human Neuroscience*, *15*(August), 1–24. https://doi.org/10.3389/fnhum.2021.702520

Todorovic, A., & de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields.

Journal of Neuroscience, *32*(39), 13389–13395. https://doi.org/10.1523/JNEUROSCI.2227-12.2012

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*(3), 434–464. https://doi.org/10.1111/j.1551-6709.2009.01077.x

Tronstad, C., Amini, M., Bach, D. R., & Martinsen, Ø. G. (2022). Current trends and opportunities in the methodology of electrodermal activity measurement. *Physiological Measurement*, *43*(2), 02TR01. https://doi.org/10.1088/1361-6579/ac5007

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945. https://doi.org/10.1162/jocn.2009.21131

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit Perceptual Anticipation Triggered by Statistical Learning. *Journal of Neuroscience*, *30*(33), 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010

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

Veale, R., Hafed, Z. M., & Yoshida, M. (2017). How is visual salience computed in the brain? Insights from behaviour, neurobiology and modeling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714). https://doi.org/10.1098/rstb.2016.0113

Vinberg, K., Rosén, J., Kastrati, G., & Ahs, F. (2022). Whole brain correlates of individual differences in skin conductance responses during discriminative fear conditioning to social cues. *eLife*, *11*, e69686. https://doi.org/10.7554/eLife.69686

von Kriegstein, K., Smith, D. R. R., Patterson, R. D., Ives, D. T., & Griffiths, T. D. (2007). Neural Representation of Auditory Size in the Human Voice and in Sounds from Other Resonant Sources. *Current Biology*, *17*(13), 1123–1128. https://doi.org/10.1016/j.cub.2007.05.061

von Kriegstein, K., Warren, J. D., Ives, D. T., Patterson, R. D., & Griffiths, T. D. (2006). Processing the acoustic effect of size in speech sounds. *NeuroImage*, *32*(1), 368–375. https://doi.org/10.1016/j.neuroimage.2006.02.045

Vroomen, J., & Gelder, B. de. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1583–1590. https://doi.org/10.1037/0096-1523.26.5.1583

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *Journal of Neuroscience*, *32*(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Wacongne, C., Labyt, E., Van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20754–20759. https://doi.org/10.1073/pnas.1117807108

Wang, C.-A., Boehnke, S. E., Itti, L., & Munoz, D. P. (2014). Transient pupil response is modulated by contrast-based saliency. *Journal of Neuroscience*, *34*(2), 408–417. https://doi.org/10.1523/JNEUROSCI.3550-13.2014

Wang, C.-A., & Munoz, D. P. (2015). A circuit for pupil orienting responses: Implications for cognitive modulation of pupil size. *Current Opinion in Neurobiology*, *33*, 134–140. https://doi.org/10.1016/j.conb.2015.03.018

Wang, C.-A., & Munoz, D. P. (2021). Differentiating global luminance, arousal and cognitive signals on pupil size and microsaccades. *The European Journal of Neuroscience*, *54*(10), 7560–7574. https://doi.org/10.1111/ejn.15508

Wang, C.-A., White, B., & Munoz, D. P. (2022). Pupil-linked Arousal Signals in the Midbrain Superior Colliculus. *Journal of Cognitive Neuroscience*, *34*(8), 1340–1354. https://doi.org/10.1162/jocn_a_01863

Wang, Y. C., & Egner, T. (2023). Target detection does not influence temporal memory. *Attention, Perception, & Psychophysics*, *85*(6), 1936–1948. https://doi.org/10.3758/s13414-023-02723-3

Waschke, L., Tune, S., & Obleser, J. (2019). Local cortical desynchronization and pupil-linked arousal differentially shape brain states for optimal sensory performance. *eLife*, *8*, e51501. https://doi.org/10.7554/eLife.51501

Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (2012). How Each Prosodic Boundary Cue Matters: Evidence from German Infants. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00580

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wessa, M., & Flor, H. (2007). Failure of Extinction of Fear Responses in Posttraumatic Stress Disorder: Evidence From Second-Order Conditioning. *American Journal of Psychiatry*, *164*(11), 1684–1692. https://doi.org/10.1176/appi.ajp.2007.07030525

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, *8*(3), 229–256. https://doi.org/10.1007/BF00992696

Willmore, B. D. B., & King, A. J. (2023). Adaptation in auditory processing. *Physiological Reviews*, *103*(2), 1025–1058. https://doi.org/10.1152/physrev.00011.2022

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2010). Bayesian online learning of the hazard rate in change-point problems. *Neural Computation*, *22*(9), 2452–2476. https://doi.org/10.1162/NECO_a_00007

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2013). A Mixture of Delta-Rules Approximation to Bayesian Inference in Change-Point Problems. *PLOS Computational Biology*, *9*(7), e1003150. https://doi.org/10.1371/journal.pcbi.1003150

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, *21*(3–4), 147–163. https://doi.org/10.1027/0269-8803.21.34.147

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*(12), 532–540. https://doi.org/10.1016/j.tics.2009.09.003

Woods, K. J. P., Sampaio, G., James, T., Przysinda, E., Hewett, A., Spencer, A. E., Morillon, B., & Loui, P. (2024). Rapid modulation in music supports attention in listeners with attentional difficulties. *Communications Biology*, *7*(1), 1376. https://doi.org/10.1038/s42003-024-07026-3

Yarden, T. S., Mizrahi, A., & Nelken, I. (2022). Context-Dependent Inhibitory Control of Stimulus-Specific Adaptation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *42*(23), 4629–4651. https://doi.org/10.1523/JNEUROSCI.0988-21.2022

Yon, D., & Frith, C. D. (2021). Precision and the Bayesian brain. *Current Biology*, *31*(17), R1026–R1032. https://doi.org/10.1016/j.cub.2021.07.044

Yon, D., Gilbert, S. J., de Lange, F. P., & Press, C. (2018). Action sharpens sensory representations of expected outcomes. *Nature Communications*, *9*(1), 4288. https://doi.org/10.1038/s41467-018-06752-7

Yu, A. J. (2012). Change is in the eye of the beholder. *Nature Neuroscience*, *15*(7), 933–935. https://doi.org/10.1038/nn.3150

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event Perception: A Mind/Brain Perspective. *Psychological Bulletin*, *133*(2), 273. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., & Swallow, K. M. (2007). EVENT SEGMENTATION. *Current Directions in Psychological Science*, *16*(2), 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology. General*, *130*(1), 29–58. https://doi.org/10.1037/0096-3445.130.1.29

Zeki, S., & Shipp, S. (1988). The functional logic of cortical connections. *Nature*, *335*(6188), 311–317. https://doi.org/10.1038/335311a0

Zénon, A., & Krauzlis, R. J. (2012). Attention deficits without cortical neuronal deficits. *Nature*, *489*(7416), 434–437. https://doi.org/10.1038/nature11497

Zhang, S., Hu, S., Chao, H. H., Luo, X., Farr, O. M., & Li, C. R. (2012). Cerebral correlates of skin conductance responses in a cognitive task. *NeuroImage*, *62*(3), 1489–1498. https://doi.org/10.1016/j.neuroimage.2012.05.036

Zhao, S., Chait, M., Dick, F., Dayan, P., Furukawa, S., & Liao, H.-I. (2019). Pupil-linked phasic arousal evoked by violation but not emergence of regularity within rapid sound sequences. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-12048-1

Zhao, S., Skerritt-Davis, B., Elhilali, M., Dick, F., & Chait, M. (2025). Sustained EEG responses to rapidly unfolding stochastic sounds reflect Bayesian inferred reliability tracking. *Progress in Neurobiology*, *244*, 102696. https://doi.org/10.1016/j.pneurobio.2024.102696

Zhao, S., Wai Yum, N., Benjamin, L., Benhamou, E., Yoneya, M., Furukawa, S., Dick, F., Slaney, M., & Chait, M. (2019). Rapid ocular responses are modulated by bottom-up driven auditory salience. *The Journal of Neuroscience*, *39*(39), 7703–7714. https://doi.org/10.1523/jneurosci.0776-19.2019

Author Contribution

Chapter 2: The author was responsible for the study and stimulus design, data acquisition, data analysis, and writing of the submitted manuscript. Maria Chait contributed to the study and stimulus design, writing of the submitted manuscript, and provided guidance on data interpretation. Roberta Bianco, Marcus Pearce, and Edward Hall provided advice in data interpretation and refined the submitted manuscript.

Chapter 3: The author was responsible for the study and stimulus design, data acquisition, data analysis, and writing of the manuscript. Maria Chait contributed to the study and stimulus design, provided advice on data interpretation and contributed to manuscript revision. Marcus Pearce provided guidance on model implementation and interpretation.

Chapter 4: The author was responsible for the study and stimulus design, data acquisition, data analysis, and writing of the manuscript. Maria Chait contributed to the study and stimulus design, provided advice on data interpretation and contributed to manuscript revision.

Appendix chapter: The author was responsible for the study and stimulus design, data acquisition, data analysis, and writing of the manuscript. Maria Chait contributed to the study and stimulus design, provided advice on data interpretation and contributed to manuscript revision. Toshio Irino contributed to the stimulus design and provided guidance on vocoder implementation.