Brain Responses Track Patterns in Sound

Rosemary Southwell

Thesis submitted in partial fulfilment for the degree of Doctor of Philosophy.

Ear Institute, Faculty of Brain Sciences, University College London, UK.

May 2019

Supervisors:

Professor Maria Chait

Professor Karl Friston

Declaration

I, Rosemary Southwell, 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.

Rosemary Southwell

August 2019

Abstract

This thesis uses specifically structured sound sequences, with electroencephalography (EEG) recording and behavioural tasks, to understand how the brain forms and updates a model of the auditory world. Experimental chapters 3-7 address different effects arising from statistical predictability, stimulus repetition and surprise. Stimuli comprised tone sequences, with frequencies varying in regular or random patterns. In Chapter 3, EEG data demonstrate fast recognition of predictable patterns, shown by an increase in responses to regular relative to random sequences. Behavioural experiments investigate attentional capture by stimulus structure, suggesting that regular sequences are easier to ignore.

Responses to repetitive stimulation generally exhibit suppression, thought to form a building block of regularity learning. However, the patterns used in this thesis show the opposite effect, where predictable patterns show a strongly enhanced brain response, compared to frequency-matched random sequences. Chapter 4 presents a study which reconciles auditory sequence predictability and repetition in a single paradigm. Results indicate a system for automatic predictability monitoring which is distinct from, but concurrent with, repetition suppression.

The brain's internal model can be investigated via the response to rule violations. Chapters 5 and 6 present behavioural and EEG experiments where violations are inserted in the sequences. Outlier tones within regular sequences evoked a larger response than matched outliers in random sequences. However, this effect was not present when the violation comprised a silent gap.

Chapter 7 concerns the ability of the brain to update an existing model. Regular patterns transitioned to a different rule, keeping the frequency content constant. Responses show a period of adjustment to the rule change, followed by a return to tracking the predictability of the sequence.

These findings are consistent with the notion that the brain continually maintains a detailed representation of ongoing sensory input and that this representation shapes the processing of incoming information.

Impact Statement

My research investigates behavioural and brain responses to abstract sound sequences in young human volunteers with no known hearing or neurological problems. Automatic brain responses to sounds are a widely-used measure in clinical neurological research. The way the brain responds to sound is known to be impacted in a range of conditions, such as schizophrenia, autism, disorders of consciousness and even in (otherwise asymptomatic) mild traumatic brain injury. However, there are still some crucial elements we do not fully understand about how the brain extracts patterns from sensory input even in the non-clinical population; and this is what my research focuses on.

It is thought that our perception comes from the interaction between incoming sensory information and the brain's current 'best guess' of the state of the world. An important aspect of brain function in general is in separating important 'signal' from irrelevant 'noise'. In hearing, this means finding the balance between what our ears seem to tell us and what we already know about the world. Making good use of such prior knowledge requires that the brain always tracks the wider *context* within which every momentary sensation is embedded.

Previous research from our lab has shown that regular patterns are rapidly detected, even when a listener's attention is engaged elsewhere; as revealed by a larger brain response whilst a pattern is occurring. The size of the brain response appears to closely track the degree of predictability, as revealed by using different types of pattern. This is potentially one such mechanism which allows the brain to use context to optimally learn about the world.

I envisage this work will disentangle the roles of several possible mechanisms or how the brain learns about regularities, which have been difficult to dissociate in previous studies. For instance, passive neurophysiological effects versus expectation-driven, 'top-down' influences. My project will contribute to better understanding the neural mechanisms for detecting pattern and using this to reason about the world. This basic research is vital to inform future research into the underlying neural basis of altered information processing in various medical conditions, for instance psychosis and auditory processing disorders.

Acknowledgements

To study at UCL over the past four-plus years has been an incredibly enriching experience. I am grateful to have been so generously welcomed into such a dense concentration of brilliant minds, bathed in the glow of the exciting ideas shared between them. I would especially like to thank Karl Friston and the theoretical neurobiology group and Torsten Baldeweg for edifying discussions during pivotal moments.

Sven Bestmann and Dimitris Pinotsis welcomed me graciously into their research projects in my rotation year, during which I learnt so much, before settling into my home in the Chait lab. Maria runs an incredibly inspiring research group, which I am so thankful to have been a part of. Her dedication, enthusiasm and sparky critical spirit drives us all to do our best. I thank all the lab members past and present for their support, camaraderie and brilliant ideas: Nicolas, Mathilde, Sijia, Daniel, Kate, Roberta, Alice, Alex, Ed, Ulrich, Gabi, Sami and more; not to mention the community of the Ear Institute as a whole. I would also like to thank the Masters students who I worked with on some of the projects presented in this thesis: Candida, Nicolas, Cecile and Anna; for their energising enthusiasm to learn, and their hard work.

This thesis represents but one thread in the last few years of my life, and it is only through the enrichment of other people and passions that I was capable of producing this work. I never known friendship so easy and so profound as in this London-based chapter of my life. Heartfelt thanks to Jess, Frances, Elizabeth, Clare, Louis, Andrew, Rosie, Deniz, Martin, Davine, Berenice, Mikel, Alison, and so many other good eggs.

Finally, my warmest gratitude goes to my mother Ruth, sister Evelyn and my husband Ewan; who from near and afar have always helped me to believe in myself.

Contents

List of Figures

Abbreviations

A1 – primary auditory cortex
alph – alphabet size
ANOVA – analysis of variance
BOLD – blood-oxygen-level dependent
d' – d prime
DC – direct current
DCM – dynamic causal modelling
DSS – denoising source separation
EEG – electroencephalography
Hz – Hertz
ICA – independent component analysis
ISI – inter-stimulus interval
ITI – inter-trial interval
ITPC – inter-trial phase coherence
LFP – local field potential
MEG – magnetoencephalography
MMN – mismatch negativity
MNE – minimum-norm estimate
ms – milliseconds
PCA – principle component analysis
RAND – random sequence
REG – regular sequence
RMS – root mean square
fMRI – functional magnetic resonance imaging
RP – repetition positivity
RT – reaction time
s – seconds
SEM – standard error in the mean
SNR – signal to noise ratio
SOA – stimulus onset asynchrony
SSA – stimulus-specific adaptation
RS – repetition suppression
µV – microvolts

Chapter 1. Introduction

1.1 From Sound to Perception

Often proceeding beneath the radar of our conscious awareness, the auditory system carries out a remarkable task. Sound is constantly evolving, intangible and exists only as a thread through time. Within milliseconds, ripples in pressure at the tympanic membrane are converted into spatial patterns of electrical activity in the auditory nerve. The ensuing cacophony of currents and action potentials somehow allows a percept to settle: the ripples become a voice in a crowd, a familiar tune, or a warning. Although the result feels stable, the meaning is not inherent in the input. The current best understanding of how the brain instantiates this process is that, rather than a unidirectional inward progression of data, the brain seeks to incorporate existing knowledge with information from the sensorium in a delicate balance of learning and inference. Crucial to this is the recognition of regularity in the world, for it is a deep though tautological truth that understandable phenomena possess some form of statistical predictability.

Hearing has long been a fertile ground for investigating how automatic and rapid information-processing is implemented by the brain, and has revealed much support for the view of the brain as an active hypothesis-forming organ. This evidence mostly takes the form of context-sensitive responses: where sounds which are considered acoustically identical in isolation actually reveal rather different effects when presented within a different context. One form of prior context which is often used is regularity itself, and its influence measured by proxy of the response to violations of the regularity. This thesis seeks to explore the influence of regularity more directly, using sounds that differ principally by the regularity of their ordering in time, whilst keeping constant the acoustic properties of their constituent parts. Building on recent work using similar techniques, I will present a series of experiments on human listeners, which together provide evidence that the brain tracks ongoing patterns in sounds, automatically, and this has immediate impact on contextualising responses to violations of expectations in a manner consistent with predominant quantitative and qualitative models of brain function.

1.1.1 Structure in natural sounds

Natural environmental sounds are characterised by stable statistical rules (Attias and Schreiner 1997; Singh and Theunissen 2003; Turner 2010), reflecting the regularity inherent in their underlying physical and biological causes. This has been true throughout the entire history of the evolution of the auditory system, over the past three hundred million years or so. It is therefore unsurprising that the present-day mammalian auditory system seems precisely tuned to efficiently recognise and represent regularity in sound (Lewicki 2002; Rieke et al. 1995; Nelken et al. 1999; Daikoku 2018).

Animal calls and the human voice are tonal, harmonic complexes of arithmetically-related frequencies forming a perfect, regularly spaced ladder in the frequency domain. Vocalisations are also often patterned in time, incorporating regular modulation in pitch and time, and repetition of motifs (Wiley and Richards 1978; Bradbury and Vehrencamp 1998; Berwick et al. 2011). Rhythmic sound even arises from the temporally-patterned nature of animal locomotion (Delcomyn 1980).

Speech, perhaps the most fundamental sonic activity in our lives, consists of nested regularities occurring over multiple timescales (Rosen 1992). Fine-grained patterns in both timing and frequency information are crucial to discerning it (e.g. Popham et al. 2018). Music is governed by implicitly-learned abstract rules, and the making and breaking of these rules is fundamentally what makes music so compelling to us (Gebauer et al. 2012; Zatorre and Salimpoor 2013). In today's technological era, we bring ever-more precisely regular sounds into our own environments. Our daily lives are increasingly filled with countless stereotyped noises from the user-interfaces of our devices and vehicles.

1.1.2 Auditory scene analysis

The task of the hearing brain is to quickly assess the most likely sources of sound, separate that which is most behaviourally pertinent from other sounds, and follow it through time. In the real world, sound waves arriving at the ears are the result of multiple overlapping causes, degraded by multiple sources of noise. The archetypal example, employed by auditory literature, of an ecologically-valid situation that *Homo sapiens* often finds herself in is the cocktail party (Cherry 1953; McDermott 2009). The voice of the person to whom she is listening is jumbled with a multitude of sound sources including other voices, music and clattering crockery. The resultant waveform arriving at each ear is fundamentally ambiguous: mathematically speaking, an infinite array of possibilities exist as to possible components that would sum to produce it: this is an ill-posed problem (Hadamard 1923). We can however (usually) disentangle the sound sources quite successfully: with recourse to knowledge from other senses, from experience of the past, and from various properties of the sounds themselves.

This process is often referred to as *auditory scene analysis* (Bregman 1990), which delineates a view of auditory perception informed by the Gestalt school of psychology: a key tenet of which can roughly be translated as *"The whole is* *something else than the sum of its parts"* (Koffka 1935). In other words, auditory perception represents a form of inference (Helmholtz and Ellis 1875). Parsing the auditory scene involves making inferences from regularities. I shall review evidence for the role of regularity detection in auditory scene analysis in the following paragraphs.

1.1.2.1 The parts

Sound information arrives in the brain already separated by frequency. As sound waves enter the cochlea, the continuously-varying stiffness along the basilar membrane means that different frequencies excite certain portions of the basilar membrane, and hence preferentially activate particular populations of the hair cells which synapse with the dendrites of the auditory nerve (Pickles 2008). This forms a biological instantiation of wavelet analysis, where the signal is decomposed as the summation of multiple arbitrary sinusoidal carriers varying in their contribution over time (Daubechies 1992). This frequency separation persists as signals enter the primary auditory cortex, where the neurons are arranged tonotopically. In order to induce a percept of the original sound-making process, the right parts must be somehow reunited at some level of representation in the brain: the perceptual correlate of this is an auditory object. Grouping of the correct features to re-create the real-world causes of sensation also implies segregation of the components forming different objects.

1.1.2.2 Inferring the whole

Similarity of auditory features between sound components tends to induce grouping into the same auditory object, and differences tend to induce segregation. This phenomenon has often been explored through the streaming paradigm, where tone pips at two (or more) different frequencies are played, each frequency repeating regularly (Bregman and Campbell 1971). This stimulus is perceptually bi-stable, meaning that the percept fluctuates between a united whole and separate parts (Denham and Winkler 2006), making it an important paradigm to study auditory scene analysis. The parameters of the stimuli affect the relative prevalence of an integrated or a segregated percept. The more distant the frequencies, and the faster the rate of tone presentation, the more likely the subjective percept is of two independent streams (Bregman and Campbell 1971). This led Bregman to hypothesise that the pitch spacing was crucial: tones that were similar in frequency tended to be grouped together in the same stream, whilst a relatively large frequency gap separating tones tended to form the boundary between two perceived streams (Bregman and Campbell 1971; Bregman 1990). As well as pitch, similarity in other sound features such as timbre (Deike et al. 2004) and amplitude envelope (Cusack and Roberts 2004) have also been shown to act as cues for co-segregation of sounds sharing the features into a single auditory object.

At the level of the auditory pathway, this has been hypothesised to be due to separate populations of feature-selective neurons responding to the putative scenes (van Noorden 1975; Fishman et al. 2001; Carlyon 2004; Fishman et al. 2004). However this is not the whole story: sounds overlapping in pitch can segregate (Vliegen and Oxenham 1999; Grimault et al. 2002) and conversely sounds distant in some feature-dimension, and coded for in distinct neural populations, can be integrated into a single object (Deutsch 1974; Elhilali et al. 2009). For instance, In the absence of spectral cues, similarity in timing of onset (and offset) can induce formation of auditory objects (Elhilali et al. 2009).

1.1.2.3 The role of spectro-temporal patterns

In reality, competing sounds in the cocktail party will partly overlap in frequency, time and other attributes, yet are still separable. One solution to this problem is for the auditory system to detect patterns expressed in a complex dance of time and frequency, when either alone do not suffice.

Bendixen et al. (2010) used a streaming paradigm where the individual streams themselves could contain a regular pattern in frequency or intensity. They found that regularity in one of these features could increase the chance of a segregated percept. Andreou & Chait (2010) observed that temporal regularity of scene components influences the ability of the listener to segregate a target stream from an irrelevant background stream: the more regular the distractor stream, the easier it was to follow the target stream. Therefore regularity also plays an important role in segregating auditory objects, whether the regular object is specifically attended, or actively ignored, or neither.

Regularity in its simplest form, repetition, can be exploited by the auditory system to extract novel auditory objects from background noise. Kidd et al. (1994; 2003) used artificial auditory scenes composed of simultaneous tones at multiple, randomly-changing frequencies, and showed that a regularly-repeating target stands out as a separate stream, so long as the background changed whilst the target tone remained constant. This effect remains when the target ‘figure’ is composed of multiple frequencies across a wide band, overlapping with the random background, and even varying in frequency over time (Teki et al. 2013; 2016). Extending this idea to naturalistic stimuli, McDermott et al. (2011) measured the ability of listeners to segregate a synthesised, repeated target from entirely spectro-temporally overlapping noise, and found that so long as the noise varied on each presentation of the mixture, the target could be recovered, even when attention was engaged elsewhere (Masutomi et al. 2016). In all these experiments, repetition enabled streaming of a hitherto novel sound. This demonstrates that the auditory system is capable of rapidly recognising abstract pattern anew based on serial repetition of a spectro-temporal regularity.

1.1.2.4 Predictions are key to auditory scene analysis

The recognition of temporal regularity allows the brain to form valid predictions about how an auditory object unfolds over time, and it is thought that such predictions indeed play a crucial role in auditory scene analysis (Denham and Winkler 2006; Winkler et al. 2009; 2012; Winkler and Schröger 2015). A predictive rule, which can account for a subset of sound components over time, leads to these components being grouped as a perceptual whole (for a review see Winkler et al. 2009; Andreou and Chait 2010; e.g. Bendixen et al. 2010; 2013). Importantly, multiple predictive schemes can be somewhat compatible with the input. The perception of ambiguous stimuli can be stabilised by prior acoustic events which render one interpretation more parsimonious (e.g. Chambers et al. 2017). Taking into account such prior information, the predictive scheme that best explains the incoming patterns of stimulation dominates later perception and action (Winkler et al. 2005). So long as the properties of the learnt rule remain relatively stable, these expectations will usually be met. Expected events are generally processed faster and more accurately (de Lange et al. 2018; Rimmele et al. 2018).

Denham and Winkler (2006) argue that the same predictive processing lies at the core of following existing sources (such as segregating voices in a conversation) and the detection of change (such as a familiar friend’s voice entering the room). When expectations arising from the current sensory hypothesis are not met consistently, this forms the basis for detecting a change, and potentially diverting processing resources to learn a new predictive rule. In fact, this process of forming and refining predictions about sensory input is now thought to be fundamental to brain function, and will be discussed in more detail in §1.3. Knowledge supporting the formation of predictions can arise from the recent history of the stimulus itself, or can be drawn from past experience over a lifetime (such as in the case of knowledge of musical rules and language), both of which Bregman (1990) refers to as schema-based segregation. However, this thesis investigates the acquisition of rules from novel, abstract stimuli with no such cultural or linguistic context.

We have seen in this section that spectro-temporal patterns describe the sounds elicited by natural causes in the world, and that the detection of spectro-temporal patterns is a crucial component of auditory scene analysis, by allowing the formations of hypotheses about future incoming information. The following paragraphs outline the evidence for regularity-sensitive auditory processing, in the form of experimental work directly measuring neural evoked responses in humans and animals.

1.2 Regularity-Sensitive Neural Responses

1.2.1 Auditory-evoked responses

1.2.1.1 Direct neuronal recordings

The experiments in this thesis use scalp electroencephalography (EEG) to provide insight into the regularity-detection mechanisms in the human brain. However, context-sensitivity is present in the very earliest of subcortical responses, as measured using direct neuronal recordings.

The term adaptation is often used to describe reduction in responses to repeated stimuli, but for clarity this thesis uses the term repetition suppression to refer to the former. ‘Adaptation’ will refer specifically to any ‘passive’ habituation of neurons to repeated stimuli caused by fatigue, for example of neuronal firing rate or through synaptic depression or long-term depression (see Grill-Spector et al. 2006 for a review). Even this form of adaptation is a powerful mechanism for representing statistical structure in the world (Fairhall et al. 2001). Stimulus-specific adaptation (SSA) refers to the tendency of single neurons to respond to the same stimulus with greater magnitude when it is presented rarely than commonly, whilst responsivity is preserved to other stimuli to which the neuron is sensitive. SSA has been observed in primary auditory cortex (A1) of the cat (Ulanovsky et al. 2003) and in even earlier processing stages: in the thalamus (Antunes et al. 2010) and the inferior colliculus (Malmierca et al. 2009). SSA can account for the encoding of stimulus history over multiple timescales, up to possibly hundreds of seconds (Ulanovsky et al. 2004).

Automatic regularity sensitivity goes beyond SSA; representing a more nuanced memory trace already at the level of the midbrain. For instance, a subset of neurons in the inferior colliculus in rats are sensitive to the spectro-temporal patterning in sound sequences (Malmierca et al. 2019). Using similar spectro-temporally patterned sequences to this thesis, Barcazk et al. (2018) found the macaque thalamus to be sensitive to sequential structure.

Regularity-sensitive modulations in single-neuron activity form the basis of the pattern detection capacities which we can measure through brain imaging and behaviour, reviewed below. Where relevant, known equivalences between responses from single-unit and human neuroimaging are stated.

1.2.1.2 Onset responses in EEG

Sound onset evokes a stereotyped series of response components, which can be measured using EEG and MEG. Within the first fifty or so milliseconds, the feedforward sweep of initial auditory processing is revealed. Successive components originate first from subcortical sources, then increasingly from primary auditory generators in Heschl's Gyrus (HG), and later from wider sources in supratemporal cortex (Picton et al. 1974; Godey et al. 2001; Yvert et al. 2005). For scalp-measured EEG, such as used in this thesis, these early components, termed the middle-latency response, or MLR, generally appear as an single unresolved peak, with a positive polarity over the centre of the scalp, known as the P1 (or P50) (Liégeois-Chauvel et al. 1994). At around 100 ms, EEG responses show a prominent peak with negative polarity over fronto-central electrodes; this is the N1 (referred to as the N1m for MEG, hereafter N1 refers to both). These onset responses occur whether sounds are attended or unattended, and even in the absence of consciousness (Weitzman and Kremen 1965). Nevertheless, there is evidence that even such early responses are sensitive to contextual regularities, and as such are relevant for discussing the process of initial regularity extraction.

The N1 amplitude and latency depend on the sound's physical parameters such as pitch, loudness and envelope (Beagley and Knight 1967; Picton et al. 1978; Roberts et al. 2000; Billings et al. 2017). Additionally, the same sound can produce different onset responses dependent on preceding context. Both the N1 and the MLR show attenuation and/or delay on successive presentations of the same tone (Slabu et al. 2010; Grimm et al. 2016); an effect potentially explainable by neuronal adaptation. However, the opposite effect, an enhancement with repetition, on the P1 has also been observed (Costa-Faidella et al. 2010). Either way, this phenomenon has been interpreted as implementing a primitive form of sensory memory; forming a building-block for ever more refined representations of auditory structure; discussed in more detail in §1.2.3.

1.2.1.3 Later auditory responses reflect cognitive processes

Following this is generally another positive deflection around 200 ms, the P2, with similar EEG topography to the N1. By this latency, the response reflects multiple active sources, including feedback connections from outside the auditory cortex (Garrido et al. 2007). The P2 amplitude has been shown to reflect the build-up of regular structure in stimulus streams, showing an increase over repetitions (Costa-Faidella et al. 2010; 2011; see also §1.2.3).

Later auditory responses are more variable, visible as more indistinct peaks in the evoked response, and reflect activity of attentional networks and other cognitive processes. The P3a, at around 300 ms, is thought to accompany attentional orienting following the detection of a change in a sound (Wronka et al. 2008; Masson and Bidet-Caulet 2019), or indeed in a visual stimulus. Its sibling, the P3b, reflects subsequent processing the behavioural significance of the change (reviewed by Friedman et al. 2001).

Related to the above discussion of auditory scene analysis, EEG responses have been identified as a correlate of the streaming of an object based on spectro-temporal coherence. The first of these, the so-called object-related negativity, or ORN, occurs at around 150-300 ms, and reflects the automatic detection of a coherent auditory object. The second at around 500 ms reflects task-relevant processing of the target object (Tóth et al. 2016).

1.2.1.4 Sustained and offset responses

Following sound onset is a sustained response which persists as long as the stimulus continues, although as one progresses up the auditory hierarchy, this becomes attenuated (Phillips et al. 2002), such that measured neural responses largely reflect changes in the level of stimulation rather than the absolute stimulus intensity. Ongoing sounds which are also changing throughout the stimulus tend to produce stronger sustained responses, though this may be due to summation of individual onset responses (Pantev et al. 1993). Sustained responses may also reflect the accumulation of evidence throughout the stimulus, such as revealed in the *contingent negative variation* (CNV; Tecce 1972; Chennu et al. 2013). Sound offset also evokes a measurable response; although this is less prominent than the onset response, it appears to be generated by similar sources (Hari et al. 1987; Phillips et al. 2002). It may be due to the cessation of stimulus-induced inhibition causing a rebound in activity (Hillyard and Picton 1978). However, Chapter 4 in this thesis will present evidence that the offset response reflects the strength of complex regularities in the preceding stimulus.

1.2.2 Deviance responses reflect regularity extraction

Patterned sound sequences, which include an eventual *deviant* sound which violates the predictive rule, have been used to obtain evidence for the sensitivity of the brain to auditory regularity. If the violation evokes a measurable response, one may infer that the pattern was recognised in the first place. There is an abundance of auditory studies which use such an electrophysiological effect, the Mismatch Negativity (MMN), to demonstrate the automatic extraction of a range of regular structures from sound sequences. Presented in the context of a repeated *standard* tone, a deviant tone which differs in its physical properties (usually pitch) elicits a greater measured neural response; this procedure is known as the *oddball* paradigm (Näätänen et al. 1978). The roving-standard paradigm (Cowan et al. 1993; adapted by Baldeweg et al. 2004; Garrido et al. 2008) is a modification of the oddball paradigm, where standards at a given frequency are repeated a variable number of times, before changing to a different standard frequency. Because the same physical stimuli act as standards and deviants, this ensures that the difference in response to standard and deviant tones (i.e. the MMN) cannot be explained purely by the fact that different frequencies were used. The first tone of the new standard frequency indeed elicits an MMN (Baldeweg et al. 2004; Näätänen et al. 2004; Haenschel et al. 2005; Costa-Faidella et al. 2011).

The MMN is not an EEG component but a difference between components: the response to the standard is subtracted from that to the deviant, and the MMN is manifest in the resulting waveform as a relative negativity over fronto-central channels, at a latency of around 100-250ms. Whilst this is broadly similar in timing and topography to the N1, many argue separate neural processes play a part in its generation, as evidenced (for example) by the presence of sources outside of auditory cortex (e.g. Alho 1995; Garrido, Kilner, Stephan, et al. 2009). The MMN is often claimed to represent the extraction of the regularity rule defined by successive repetition. Initially, this was described in terms of the formation of a memory trace (Näätänen et al. 1993) to which the deviant tone is subsequently compared. When the pattern is broken by the deviant tone, automatic change-detection processes produce the negative deflection of the MMN, in addition to the standard onset responses to the deviant tone.

1.2.2.1 The MMN is sensitive to the degree of violation

The magnitude and latency of the MMN has been shown to vary with the extent of violation in some physical feature dimension, with greater magnitude and/or decreased latency found to more extreme deviants. For instance, Amenedo & Escera (2000) found the magnitude of the MMN to vary with how different in duration the deviant sounds were to the standard sounds. The greater the loudness difference between standards and deviants, both the earlier and the larger the MMN (Näätänen et al. 2007). Similar effects have been found with respect to deviant frequency distance (Todorovic and de Lange 2012; Grotheer and Kovács 2015) although this is not always reported (Desimone 1996; Budd et al. 1998; Grill-Spector et al. 2006; Kok, Jehee, et al. 2012). In a more abstract sense, a deviant occurring after a longer train of identical standards forms a stronger violation than a deviant following only a few standards. Indeed, the MMN is sensitive to the number of preceding standards (Haenschel et al. 2005).

1.2.2.2 The MMN reflects violation of complex regularities

Other studies have used standards which themselves vary in a predictable manner, and deviants which break this pattern, to show that the MMN can occur to more abstract violations of a predictive rule. In the context of paired standard tones, which consistently ascend within the pair, a deviant descending tone pair elicits an MMN (Saarinen et al. 1992). Many other studies report mismatch responses using variations of this kind of frequency rule (Todorovic and de Lange 2012). The MMN to such abstract violations can be elicited even in situations where the listener is not attending to the stimuli and/or is unaware of the predictive rule (Grotheer and Kovács 2015). Also, in paradigms using multiple deviants with different chances of occurring, the evoked mismatch response magnitude is inversely proportional to the probability of the deviant (Feuerriegel et al. 2017), which can be interpreted as sensitivity to summary-statistic descriptions of regularity. A similar effect is seen in auditory cortex of cats (Grotheer and Kovács 2016), where the less probable a deviant tone, the greater the response of single neurons.

Beyond tone-pip sequences, naturalistic paradigms also show that the MMN reflects rule-based processing. In musical sequences, even when unattended, out-of-key and out-of-tune notes elicit a larger MMN than in-key notes (Segaert et al. 2013). Detecting such deviants requires implicit knowledge of musical structure, so this demonstrates that early auditory responses are sensitive to abstract culturally-acquired regularity; however it may reflect subtly different neuronal processing than, say, the oddball MMN (e.g. Petit et al. 2006; Henson et al. 2008). Even linguistic violations elicit an MMN; the latency of which depends on the experience of the listener; it being earlier for native speakers of the language (Stefanics et al. 2018).

1.2.3 Repetition suppression as regularity detection

As inferred via the MMN, auditory-cortical responses are sensitive to the context established by regularities in preceding stimuli. One key consideration when interpreting this evidence is that the MMN is not an evoked response, rather a difference between evoked responses. Therefore, the MMN effects described so far implicitly rest on the response to the standard tones themselves, and it is important to establish whether the pattern of results is driven by the response to the deviant or to the standard. It is also important to look at how the responses to standards change over time, as this provides a more direct view of the evoked responses accompanying pattern detection.

Indeed, the N1 response to standard tones diminishes in magnitude with increasing repetitions the standard (Recasens et al. 2015). This repetition suppression explains the existence of the MMN in the oddball paradigm and casts some doubt on the existence of the MMN as an independent component of cortical responses (Friston 2005; Feldman and Friston 2010). It also explains the gradated modulation of MMN magnitude dependent on the number of preceding standards (Friston 2005; Auksztulewicz and Friston 2016).

Repetition suppression is suggested to be a correlate of the stimulus-specific adaptation (SSA) effect measured at the level of single neurons and local field potentials (e.g. Haenschel et al. 2005). Costa-Faidella et al. (2010) modelled EEG responses in a modified oddball paradigm, and showed that the time-course of adaptation of the response around the latency of the MMN over stimulus presentations resembles that seen for SSA (Ulanovsky et al. 2004; Costa-Faidella et al. 2010). However, the existence of similar repetition suppression effects in far earlier EEG components, such as the MLR, suggests adaptation in the MLR, rather than the N1 suppression, to be the most likely candidates to coincide with the SSA effect (Malmierca et al. 2014; Grimm et al. 2016).

1.2.3.1 Repetition positivity

Using the roving-standard paradigm, Haenschel et al. (2005) found a reduction in the N1 peak with successive repetitions of the standard. They observed an additional effect: a superimposed, slow positive component which increased in magnitude over repetitions, and extended from 50ms to 250 ms post-onset (Repetition positivity; see also Baldeweg et al. 2004). This positive shift served to decrease the N1 magnitude, but also to increase the P1 and P2. The repetition positivity was interpreted as a direct correlate of the encoding of the repetition rule's strength. Together, these modulations of the response to the standard fully explained the context effects on the MMN.

However, repetition positivity does not appear to be entirely consistent with a purely adaptation-based explanation. Costa-Faidella et al. (2011) show the repetition positivity is reduced in the roving paradigm when the timing of the stimulus stream is unpredictable, as compared to the usual isochronous condition. This suggests that the repetition effects in fact reflects active suppression of responses to tones enabled by their predictable temporal and spectral properties (discussed further in §1.3.3).

1.2.3.2 Interpretations of MMN revisited

Taken together, the MMN results reveal similar properties of error responses to violations of many types of auditory rules. Furthermore, mismatch responses are modulated such that their magnitude correlates with the strength of the preceding regularity rule. This rich body of work provides a valuable window onto the types of pattern which can be automatically extracted.

However, viewing similar experimental paradigms from the viewpoint of suppression of the standard tone, suggests an alternative, adaptation-based explanation which can account for some of the properties of the MMN. As evidenced by repetition suppression results described above, the MMN in oddball paradigms can be explained in terms of a suppressed N1 response, rather than the activation of a separate deviance detection mechanism (Jääskeläinen et al. 2004; May and Tiitinen 2010). This would suggest that the extraction of regularity rules does not necessarily require the formation of a predictive scheme such as proposed to account for behavioural phenomena in auditory scene analysis (§1.1.2).

Nevertheless, the MMN, and repetition suppression, cannot always be explained entirely as an adaptation-driven phenomenon. Numerous studies support the interpretation of the MMN as reflecting a more complex form of change detection, separable from modulation of basic auditory evoked responses; in particular the N1 (for a review see Näätänen et al. 2007).

Findings from the various MMN paradigms, and the different interpretations thereof, have more recently been unified with recourse to the predictive coding framework (Garrido, Kilner, Stephan, et al. 2009; Winkler et al. 2009). Predictive coding, and the evidence for its operation in the auditory system, will be discussed in more detail in §1.3. To briefly summarise, during the presentation of the standards, the auditory system forms an expectation that the stream of tone-pips will continue to repeat (or to follow whatever pattern). This extraction of the regularity rule allows the formation of predictions about what sound will come next; regardless of whether this is described as the establishment of a 'memory-trace', or is mechanistically explainable by adaptation to the standards. Either way, the relatively-enhanced deviant response results from an interaction between prior knowledge of regularities acquired during the stimulus sequence, and any discrepancy with incoming stimuli. The discrepancy, or *prediction error*, is signalled by an increase in neural activity passed up the auditory hierarchy, measurable as the MMN.

Computational modelling of cortical network activity, based on evoked responses in the roving paradigm, show that a combination of adaptation and plasticity changes in hierarchical connections, manifest over different latencies, explains the combination of repetition suppression and deviant response enhancement effects (Garrido, Kilner, Kiebel, et al. 2009). Overall, the experimental evidence indicates multiple processes of regularity extraction, progressing from the detection of simple repetitions within tens of milliseconds, to more complex rule extraction after 100 ms, revealed by the MMN to abstract regularities (as reviewed by Escera and Malmierca 2014).

1.3 Perception as Probabilistic Inference

We will return for a moment to discussing predominant theoretical perspectives on brain function, which form part of the motivation for the experiments carried out in this thesis. Then, evidence arising from auditory studies for the validity of such theories will be (re)visited.

A unifying principle in sensory perception, garnering increasing empirical support, is that of perception as inference (Helmholtz and Ellis 1875; Gregory 1980); wherein the state of the external world is deduced through combining prior knowledge with input from sensory systems. Cognition has strong parallels with the scientific process (Gregory 1980). Hypotheses, whether scientific or cognitive, represent an attempt to generalise from limited observations. This generalisation could take the form of inference about the future, but not necessarily so: a hypothesis could also concern the cause of some noisy input in the past, or interpolation from a small number of observations. There are multiple flavours of such predictive processing models falling into this overarching family, with various degrees of specificity and different neuronal implementations (Clark 2015; Spratling 2016). The key unifying characteristics are that neural activity and connectivity instantiate a generative model of the world, from which predictions are formed and compared to sensory data. Predictive coding in the brain equates to comparing an internally-generated prediction of sensory input with the actual afferent neural activity, and only communicating the difference, prediction error, in efferent activity. Prediction errors, the residual input unexplained by the current model, are minimised in order to optimise the predictive model.

1.3.1 Uncertainty in the Bayesian brain

Extending the observation of Gregory (1980) that perceptual inference resembles the scientific process of forming and testing hypotheses, is the idea of brain function as Bayesian inference.

Bayesian statistics concerns the amalgamation of present knowledge with new information optimally under uncertainty. Mathematically, Bayesian inference describes how prior information and observations are combined to estimate the posterior probability of an underlying cause, such as an auditory object corresponding to a listener’s phone ringing in the milieu of a noisy party. In the following equation, the probability of a candidate cause of sensory input is estimated as proportional to the *prior* probability of the cause occurring in general (how often does my phone ring?) multiplied by the probability that the candidate cause would generate this exact pattern of input (what is the probability that my phone ringing would cause these exact sounds?); this latter term is the *likelihood*. The estimated probability of the cause *given* the sensory input is termed the *posterior*; this is the strength of the final belief that the phone is indeed ringing.

$$P(cause|input) \propto P(input|cause)P(cause)$$

The brain may entertain multiple possible causes in such a way, selecting the cause with the greatest posterior probability to drive subjective perception and behaviour.

This mathematical framework has parallels with phenomena in perception and action, and has inspired a large body of experimental work, testing quantitative and qualitative predictions arising from the hypothesis that the brain performs Bayesian inference. A frequent approach is to show that a behaviour reflects a Bayes-optimal combination of 'top-down' expectation and 'bottom-up' input, given controlled manipulation of these quantities by the experimenter (reviewed by Knill and Pouget 2004).

The key point of relevance to the present thesis is that uncertainty plays a crucial role in optimal inference, as well as in perception (Yu and Dayan 2005; Bach and Dolan 2012). For Bayesian inference, the probabilities in the above equation can actually be represented not as single, point estimates but as probability distributions, parameterised (for instance) by their mean and variance (Knill and Pouget 2004). The variance of these distributions influences the power balance between the existing predictions and the new input, with the more precise (low variance) term having the strongest influence on the posterior belief, and hence on updating the brain's internal model (Yu and Dayan 2005). Relatedly, the tracking of different sources of uncertainty is important for determining the optimal learning rate (Courville et al. 2006). The level of noise in sensory inputs is often stable over some context, such as when speaking over a poor phone connection in a loud office; the resulting uncertainty is often called expected uncertainty. This is important for ensuring the robustness of perception or behaviour to chance fluctuations. In contrast, unexpected uncertainty refers to a truly surprising change in input, and may signal a change in contextual regularities, which, normatively, should trigger an increase in the learning rate to update the internal model (Yu and Dayan 2005). Therefore the inferential brain needs to be able to estimate and represent current uncertainty levels.

Bayesian inference and predictive coding are often discussed together, however it is important to note that the two concepts are not equivalent: Bayesian inference describes the prediction-generating process which *may* be approximately implemented neuronally using hierarchical predictive coding (Aitchison and Lengyel 2017). Predictive coding, at its heart, is a scheme for representing information parsimoniously, and doesn't itself stipulate that predictions and prediction errors are combined according to Bayes theorem.

1.3.2 Hierarchical predictive coding

Hierarchical predictive coding, where predictions are formed at higher-order brain regions about expected neural activity in lower-order regions, accounts for 'non-classical' receptive fields in visual processing (Rao and Ballard 1999). In this case, neuronal activity depends on the surrounding context in visual space, as well as on the configuration of inputs falling into the neuron's own 'classical' receptive field. Each area of the brain receives predictions from, and sends prediction errors back to, a higher level of the hierarchy. At each subsequent stage of processing, the predictions and prediction errors concern an ever more abstract and integrative representation of states in the world. This system is agnostic to whether input at a given level arrives directly from sensory neurons or as the prediction errors from another neuronal population within the brain.

1.3.2.1 Generalised predictive coding

Such a hierarchical implementation of predictive coding is at the core of Friston's (2005; 2006) free energy principle of brain function as approximate Bayesian inference. This particular formulation of predictive coding will be used to motivate hypotheses throughout the thesis, and for clarity will be referred to as generalised predictive coding (gPC) after Shipp (2016).

Friston et al. (2006) describes how a hierarchical predictive coding system implements Bayesian perceptual inference. The model parameters encode a *probabilistic* distribution, parameterised by a mean and a variance, over modelled real-world causes of sensation. The probabilistic nature of the representations allows for an approximate Bayesian inference scheme where beliefs are updated by prediction errors in proportion to the relative strength, or *precision,* of the two quantities. The priors at one hierarchical level are provided by predictions from the level above. The 'context' supplied by these prior expectations encoded by higher order brain regions can therefore influence the posterior beliefs in lower-order sensory regions. The generative model is refined by the minimisation of a statistical quantity, variational free energy, which approximates the overall discrepancy between the model and the world. Neuronally, this equates to minimising prediction error at all levels of the hierarchy in the long run. Free-energy minimisation has been shown to be a metabolically efficient path to an accurate representation of the world (Sengupta et al. 2013). Free energy minimisation also penalises complexity of the generative model, preventing overfitting of the model to empirical data; or in more concrete terms, imbues robustness of perceptual inference to noise. The free energy principle describes learning, perception and behaviour in a compelling unified theory (Friston et al. 2010).

1.3.2.2 Neuronal implementation

A plausible neuronal underpinning has been described to implement gPC. Prediction errors are passed up the hierarchy via excitatory synapses from superficial pyramidal cells, and predictions are carried in inhibitory top-down and lateral connections onto superficial from deep pyramidal cells (Bastos et al. 2012). Feedforward connections are thought to be excitatory, and feedback connections are modulatory, acting via inhibitory interneurons to suppress prediction error. Several studies have demonstrated support for this implementation (reviewed in

Bastos et al. 2012; Shipp 2016). For instance, separate neural populations represent predictions and prediction errors (Markov et al. 2013; Bell et al. 2016).

1.3.2.3 Precision & attention

Apart from prediction and prediction error, there is a third quantity which the brain must represent to allow optimal inference: precision. Precision is mathematically the inverse of variance, or in other words, of uncertainty. In gPC (as in related hierarchical Bayesian schemes; Mathys 2011) prediction errors are associated with an estimate of how certain they are, as based on inference from contextual information in much the same way as the predictions themselves. Precision wields its influence through weighting prediction errors according to their reliability; it is *precision-weighted* prediction errors which are passed up the hierarchy and used to refine the generative model. This precision parameter determines the balance of influence between prior prediction and novel sensory input, possibly through modulation of synaptic gain on prediction-error neurons, determined by modulatory connections from precision-expectation units in higher levels (Friston 2008; Shipp 2016).

Agnostic to the details of the inference scheme used by the brain, gain mechanisms have been shown to be related to selective attention (Hillyard et al. 1998; Spratling 2008), with some going further and proposing precision-weighting and attention to be equivalent (Feldman and Friston 2010; Vossel et al. 2014).

1.3.3 Predictive coding in auditory responses?

These conceptual models; that the brain instantiates a model of the world in general, and gPC in particular, can account for several phenomena in auditory responses (reviewed by Denham and Winkler 2018; Heilbron and Chait 2018).

1.3.3.1 Prediction error

According to gPC (Friston 2005) scalp-measured evoked responses predominantly reflect the prediction error signalled by superficial pyramidal cells. Predictive coding explains the auditory responses measured in MMN paradigms in terms of a mismatch between predicted and actual bottom-up neural activity; and this applies both to the standard tones and the deviants. For the standards, prediction error is suppressed in response to repeated stimuli because they are *predictable*; not because they are repeated per se. The extraction of this predictability rule by higher-order areas, through changes in synaptic connectivity, allows suppression of prediction errors arising from the sensory input (Garrido, Kilner, Stephan, et al. 2009). Modelling of evoked responses recorded in the roving paradigm indeed shows that neural responses, both to the standards and the deviants, are best explained as reflecting prediction error, and that both adaptation and regularity-learning are implicated (Garrido et al. 2008).

A recent modelling study (Rubin et al. 2016) used the single-unit recordings from Ulanovsky et al. (2004) to infer attributes of the information encoded by auditory cortex neurons. Using the assumption that neural responses in auditory cortex encode prediction error, the authors deduce properties of the predictive representation from which prediction errors are calculated. They find that models which use the recent past to calculate prediction error correlate with the measured neural responses, and models which explain the greatest proportion of neuronal activity utilise a long memory of previous sounds. Interestingly, they found most support for those models using a somewhat reduced representation of stimulus history, which constitutes a trade-off between complexity and predictive power. This study provides evidence that auditory responses to stimuli in a sequence represent prediction error, without stipulating a particular neural implementation of predictive coding.

Pieszek et al. (2013) found evoked-response correlates of both prediction and prediction errors, in an audio-visual paradigm where visual cues could also provide predictive information about tone pips. The CNV signal was interpreted as reflecting predictions, via preparation for making a behavioural response. Violations of expectations arising from auditory, visual, and bimodal predictive relationships all resulted in an increased error response. Multiple violations were associated with a linear summation of the underlying prediction errors, provide indirect evidence that multiple predictive models are maintained and evaluated simultaneously.

Beyond MMN-like paradigms, prediction errors explain responses to more complex stimuli. For instance, Blank et al. (2016), using fMRI responses to speech in noise, elegantly showed that the influence of prior expectations on multivariate patterns of neural responses were best explained via suppression of prediction error, as compared to an alternative hypothesis whereby expectations sharpen neural representations.

1.3.3.2 Expectation suppression

Interpreted in terms of predictive coding, repetition suppression such as that seen to standards in the roving oddball paradigm results from the top-down communication of predictions, acting to suppress responses to the sensory input (Auksztulewicz and Friston 2016). This position is strengthened by paradigms which manipulate expectation independently from repetition. Reduced responses to predicted stimuli in such paradigms is often termed expectation suppression. Here, stimuli which are predictable, or expected, evoke smaller responses than equivalent unpredicted stimuli, even when expectations are manipulated independently of repetition. For instance, Costa-Faidella et al. (2011) show that repetition suppression is greater when the timing of repeated tones is itself predictable. When repetitions are expected based on the context of the current block, repetition suppression of MEG responses is greater than when repetitions are infrequent in the current block (Todorovic et al. 2011).

Expectation suppression also arises when comparing identical auditory stimulation that is internally generated versus exogenously presented; in the former case, the pattern of auditory stimulation can be predicted accurately by the observer, as they were responsible for generating the stimulus (e.g. Eliades and Wang 2008; reviewed by Bendixen, SanMiguel, et al. 2012).

1.3.3.3 Anticipatory predictions

Although predictive coding doesn’t specify that the ‘predictions’ in predictive coding are necessarily *anticipatory* predictions, one of the advantages of predictive coding is its facilitation of processing input which is expected to occur at a defined future time; a situation which arises often in natural auditory scenarios.

Support for anticipatory predictions underlying auditory responses comes from the omission response (Raij et al. 1997; Hughes et al. 2001). Here, an evoked response occurs to unexpected omissions of sounds, at a similar latency to the early responses to actual sounds; often only when the preceding sequence context allows a prediction to be formed about the omitted tone’s properties (Bendixen et al. 2009). Similarly, responses to the offset of a temporally regular sequence occur with the expected timing of the next tone, given the rhythm of the preceding sequence (Andreou et al. 2015).

During sentence comprehension, the build-up of a slow negative potential correlates with the predictability of the final word in the sentence, determined by the preceding context (León-Cabrera et al. 2019). As the same potential was observed regardless of whether the sentence was presented in the auditory or visual modality, its origin is thought to be from top-down processes implementing semantic prediction.

1.3.3.4 Hierarchical auditory network

There is evidence from the auditory system in humans and other animals, as recorded both from direct neuronal recordings and at the scalp, that a hierarchically organised system of cortical and subcortical areas communicate predictions and prediction errors corresponding to increasing levels of rule complexity (see also §1.4.2).

The ‘local-global’ paradigm is a modification of the oddball paradigm, where deviant tones occur regularly at the end of every ‘chunk’ of four standard tones at a repeated frequency; such that the motif AAAAB repeats consistently. Whilst the deviant 'B' tones still elicit an enhanced response, this is reduced in blocks where the AAAAB sub-sequence repeats (Bekinschtein et al. 2009). In fact, when instead the fifth tone is another standard, i.e. a subsequence of AAAAA presented within a block of AAAAB, the expectation of a deviant occurring on every fifth tone is itself violated, a deviance response is reported. This response is later than the MMN, and has sources in a wider network including prefrontal cortex, measured with intracranial recordings (Bekinschtein et al. 2009; Chennu et al. 2013). Similar manipulations have shown different timescales of EEG responses corresponding to prediction violations evaluated at different levels of the auditory hierarchy (Grimm et al. 2011; Lecaignard et al. 2015). In a simpler paradigm contrasting expected and unexpected repetition within pairs of tones (Todorovic et al. 2011), a similar dissociation was found in evoked responses at different latencies: repetition suppression occurred around 50 ms, followed by expectation suppression between 100 and 200 ms. There is also direct evidence for hierarchical prediction errors along the auditory pathway from single-neuron recordings in rodents (Parras et al. 2017). These studies shows that expectations at multiple levels of abstraction influence evoked responses in different levels of the processing hierarchy, independently from repetition (reviewed by Escera et al. 2014).

1.3.3.5 Precision and gain

Of all the components of predictive coding, there is perhaps the least direct evidence for precision signalling in the auditory pathway. One problem is that manipulations of precision are hard to achieve independently of prediction and deviance. Another is that precision should act to increase the responses to reliable streams of input, whereas prediction should silence these responses. It appears, at least when predictability is manipulated by repetition, that the latter, suppressive effect is dominant. Using more complex spectro-temporally patterned stimuli, experiments from this lab have shown evidence for the augmentative impact of predictability on the evoked response (Barascud et al. 2016; described in more detail in Chapter 3; Sohoglu and Chait 2016a), which can be interpreted as precision-weighting. This thesis includes similar experiments, which test the relative influence of precision-weighting, prediction error and repetition on evoked responses.

1.4 Learning Sequential Regularities

Auditory predictions can be derived from many sources, including explicit and implicit visual cues, and a lifetime of previous experience with particular categories of auditory object. However, this thesis isolates a different source of auditory predictions: namely, implicit predictive relationships between subsequent elements in a sequence. Such effects are well-studied in the statistical learning literature, which uses novel and arbitrary rules to probe the initial phase of learning and evaluating predictions. Finally, putative brain areas and network schemes which are implicated in detecting and evaluating ongoing sequence structure are discussed.

1.4.1 Statistical learning

This section will review evidence from naturalistic and artificial auditory and visual statistical learning paradigms that the brain tracks statistical structure in the environment. In many ways, the findings from this body of literature are complementary to the auditory scene analysis and mismatch negativity experiments reviewed above. However, statistical learning paradigms generally probe the detection of more complex, changing and serial interdependencies than are implicated in oddball-like paradigms. Additionally, the capacity for such learning appears in different sensory modalities, in that it is capable of extracting similar transitional rules from visual, auditory and even tactile items. This ability of the brain to flexibly extract arbitrary rules is also a necessary, but not sufficient, prerequisite for corroborating both predictive processing and Bayesian accounts.

1.4.1.1 Statistical learning of transition probabilities

The field known as statistical learning provides a rich repository of studies on sequential regularity learning in particular. Humans are remarkably adept at extracting statistical interdependencies between successive elements in a sequence. This appears to be an automatic and obligatory process, as evidenced by implicit learning paradigms, where the subject is presented with rule-governed sequences whilst remaining uniformed of the existence of any structure. Artificial grammar paradigms use sequences of stimuli, governed by artificial rules expressed by the probability of transitions between successive elements in the sequence (forming a Markov chain). These sequences can be presented to subjects during an exposure phase, then *implicit* learning of the structure is probed by the responses to tokens which violate the regularities. Eight-month old infants exposed to continuous streams of artificial syllables, which are generated using such transition-probability rules, show subsequent attentional bias towards pseudo-words which would be unlikely to arise from the rules governing the sequences in the exposure phase (Saffran et al. 1996). Analogous regularities can be extracted from sequences of tone-pips (Saffran et al. 1999), numbers (Rose et al. 2005), visual forms (Turk-Browne et al. 2005) and tactile stimuli (Conway and Christiansen 2005a). Although there are similarities in statistical learning across domains, and even evidence for some shared processing, there is increasing evidence weighing in favour of somewhat encapsulated mechanisms for statistical learning in different modalities (reviewed by Milne et al. 2018). Temporal interdependencies are particularly crucial in defining sounds, and there is evidence that such statistical learning capacity is particularly biased towards the auditory modality (Conway and Christiansen 2005a).

Such sequential regularities can be used to facilitate subsequent behaviour, such as responding faster to sequence elements which follow the regularity (Rose et al. 2005; Turk-Browne et al. 2010), or mapping meaning onto the learnt pseudo-words (Graf Estes et al. 2007). This advantage exists even when participants are unaware of having learnt a rule, as assessed by de-briefing after the experiment (Rose et al. 2005; Turk-Browne et al. 2010). Behavioural facilitation afforded by statistical learning need not be limited to processing features relevant for the rule extraction. For instance, Tillmann et al. (2010) presented tone sequences generated using similar artificial grammars, whilst asking participants to judge whether target tones were in or out of tuning. Target tones that were 'grammatical' were detected as in or out of tune faster than ones which violated the grammar, even though the pitch judgement and grammaticity were orthogonal.

When a sub-sequence of elements recurs repeatedly, these elements may come to be represented in terms of 'chunks', with each chunk comprising all the stimuli in a repetition cycle (Dehaene et al. 2015). This scheme provides an alternative interpretation of the MMN findings in the local-global paradigm (§1.3.3d): the five-element sub-sequence of the type 'AAAAB' forms one chunk, and the 'deviant' AAAAA subsequence forms another. At the chunk-level, this paradigm now reduces to the basic oddball paradigm, offering an interpretation of the 'global' MMN in terms of memory representation rather than rule extraction. Such a chunking explanation can also apply to statistical learning findings such as the pseudo-word learning capacity of infants (Saffran et al. 1996). In fact, that study was explicitly designed to capture the learning of word boundaries, formed by the contrast between high-probability transitions within words, and low-probability transitions between words. Chunking and learning of transition-probabilistic rules are not mutually-exclusive: for the subject to detect regularly-occurring chunks in the first place must rest on recognising sequential relationships between elements.

1.4.1.2 Sensitivity to level of disorder in complex sequences

As described in §1.3.1, tracking the overall level of uncertainty in sensory streams is important for inference. The detection of changes in the level of disorder has been recently studied using rapid tone-pip paradigms similar to those used here. As discussed by Chait et al. (2007), transitions to a more disordered state are processed differently to changes towards order. Brain and behavioural responses to transitions to disorder are faster than symmetrically-opposite transitions to order (Chait, Poeppel, de Cheveigné, et al. 2007; Barascud et al. 2016). Recently, Zhao et al. (2018) show that pupil dilation occurs selectively to violations of regularity, and not to emergence of regularity, save in the simplest case of regular repetition of a single tone.

Sequence-level statistical properties influence subjective responses to individual elements. Agres et al. (2017) generated non-musical tone sequences, for which information-theoretic measures of expectation were estimated both at the level of individual tones and at the level of entire pseudo-melodies. In general, participants judged probe tones to be more expected the lower the information content of the tone under their perceptual model. However the strength of this relationship depended on the overall level of disorder of the sequence, being stronger in sequences governed by strong transition-probability rules, than in more entropic sequences. Sequence entropy also correlates with being harder to remember (Agres et al. 2017) and subjective perception of overall uncertainty (Hansen and Pearce 2014). Uncertainty arising from the level of entropy in the sequence may tracked separately from complexity (which follows an inverse-U relationship with respect to predictability), as evidenced by different patterns of BOLD activity parametrically varying with the two quantities (Nastase et al. 2015).

1.4.1.3 Deterministic sequence learning

In statistical learning paradigms, rules are often expressed as transition probabilities between items, with regularities embodied by a transition probability closer to 1. However, these regularities, taken to their logical extreme, become entirely deterministic. Studying the responses of the brain to entirely deterministic structure may provide an optimal window onto rule detection (a position defended in Bregman 1990). Is deterministic structure represented qualitatively differently from stochastic rules, given that perceptual inference can never be truly certain? Whilst a definitive answer to this question is beyond the scope of this thesis, both deterministic and probabilistic structure govern the sequences used herein.

Implicit learning of both deterministic and probabilistic structures is shown in serial reaction-time (SRT) tasks. Here, subjects are asked to respond specifically to each element in a temporal sequence, without being told that the sequence will repeat. Reduction in reaction times throughout the sequence is used to infer the learning process online. Post-session reports often indicate that deterministic regularities are learnt explicitly and are accessible to conscious report, whilst participants remain unaware of implicitly-learnt probabilistic rules (e.g. Vandenberghe et al. 2006). However, these findings from SRT are limited to the acquisition of slow patterns as time must be allowed for a response after each element.

In a series of experiments upon which the stimuli used in this thesis are based, deterministic structure was contrasted with probabilistic structure in rapid sequences of tone-pips (Barascud et al. 2016; see §1.5.1), which were presented to participants engaged in an incidental visual task. Regularly-repeating sequences of between 5 and 15 elements evoked an increased brain response, as compared to random sequences containing the same elements (see Figure 1.1). Here, learning of the regular structure occurred on the sub-second timescale of individual trials; assessed by behavioural detection of the emergence of regular structure; and was associated with a sustained increase in MEG and BOLD response power.

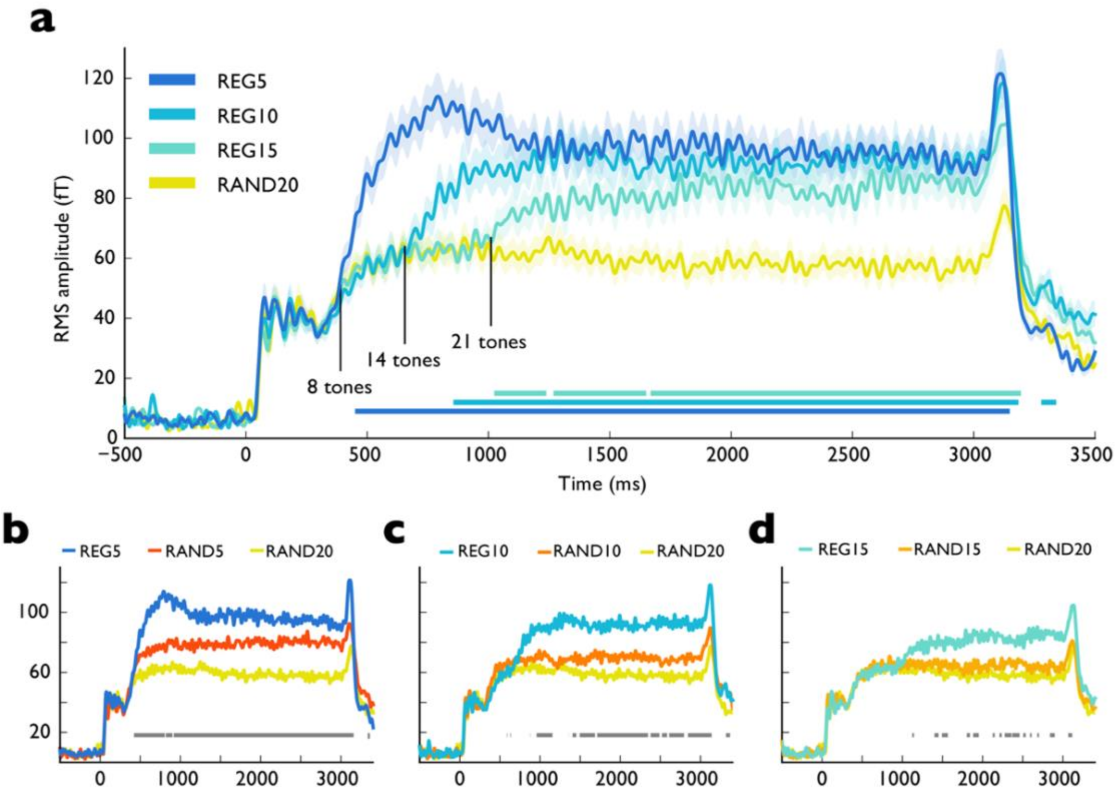

Figure 1.1 MEG responses to regular versus random sounds
Figure reproduced from Barascud et al. (2016). MEG response power to regular and random patterns of tone pips generated from different alphabet sizes. **a:** detection of the emergence of regular cycles of different lengths, as shown by a significant increase in brain response. **b, c, d:** comparison of response magnitude for regular and random patterns composed from the same alphabet of tones. Bars underneath plots denote time periods showing significant differences between pairs of conditions.

As assessed by behavioural and EEG responses to deviants, the representation of the recurring sub-sequence seems to be coded in terms of relative pitch relations, showing (some) transfer of sequence knowledge to transposed sequences (Bader et al. 2017).

1.4.2 Pattern-sensitive networks

There is converging evidence for a hierarchically-arranged network of subcortical, primary auditory, and association cortex regions involved in recognizing, tracking and evaluating auditory patterns . Depending on the acoustic properties of the stimuli, auditory scene analysis can begin in subcortical areas, though links to human perceptual measures of auditory organization are more readily established from primary and higher-order auditory cortex (reviewed by Shamma and Micheyl

2010). Inference on the basis of sequential properties seems to involve interactions between auditory and frontal regions, as well as the hippocampus.

1.4.2.1 Encoding in AC

Primary auditory cortex, and subcortical structures, extract short-term physical properties of sound elements, such as periodicity and timbre (Seger et al. 2013). As discussed in §1.2, the extraction of regularity can already be implemented by repetition suppression in A1, and even earlier stages of processing. More complex features are encoded in planum temporale (PT); outside of primary auditory cortex. PT is activated in correlation to the subjective perception of a new auditory object (Teki et al. 2011; 2016). PT also tracks entropy in sequences (Overath et al. 2007).

1.4.2.2 IFG-temporal network

The inferior frontal gyrus (IFG) plays a prominent role in domain-general sequential processing (Wang et al. 2015), as well as playing a specialised role in the perception of language (left IFG; e.g. Opitz and Friederici 2007) and pitch (predominantly right IFG ; e.g. Uluç et al. 2018).

Converging lines of evidence provide strong support for the role of frontal connectivity (specifically IFG) to auditory temporal areas in monitoring the environment for violations of regularity. fMRI, EEG, MEG and intracranial electrode measurements all find activation of this network to underlie the MMN (Doeller et al. 2003; Garrido et al. 2008; Phillips et al. 2016). Dynamic causal modelling of MEG responses allowed Phillips et al. (2015) to infer that evoked responses to auditory deviance in multiple dimensions is explained by top-down signals from inferior frontal gyrus (IFG) modulating activity in superior temporal gyrus (STG) and A1.

In the context of music, Bianco et al. (2016) find the right IFG to be a hub mediating sequencing of both auditory perception and motor responses during sequences which follow musical rules. Melodic violations are associated with increased IFG activity, in sub-regions with high connectivity to STG (Seger et al. 2013; Bianco et al. 2016), providing further evidence for a frontal-temporal monitoring of auditory pattern.

1.4.2.3 Hippocampus

The hippocampus is involved in temporally-structured memory (Jensen and Lisman 2005; Kumar et al. 2014), and may be involved in extracting adjacent dependencies in particular (Opitz and Friederici 2007). Hippocampal activation is sensitive to sequence regularity, showing relative activation to deterministic sequences with respect to random ones (Barascud et al. 2016). Interestingly, the opposite relationship between sequence predictability (indexed by entropy) and hippocampal activation was observed with non-deterministic sequences (Strange et al. 2005).

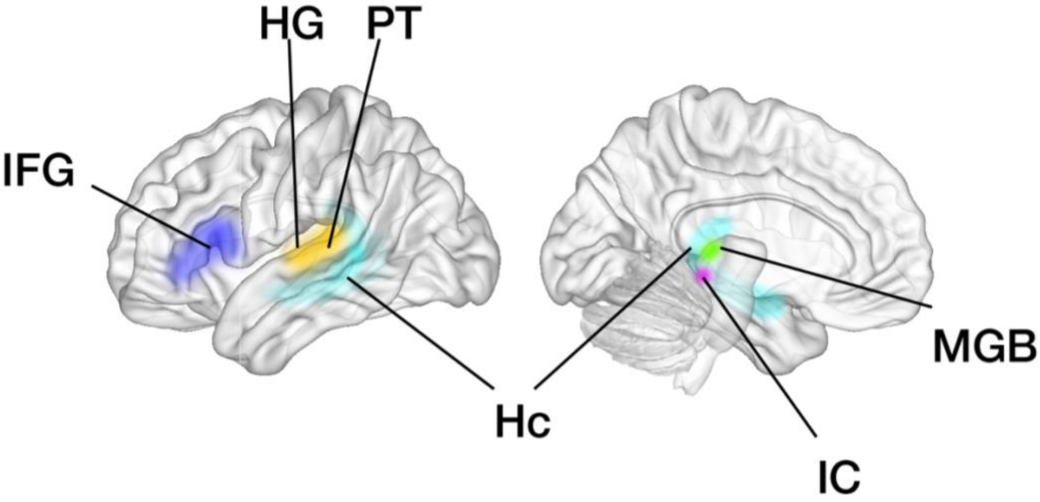

Figure 1.2 Pattern-sensitive regions
Schematic depicting the main auditory regularity-processing regions referred to in the literature and in this thesis. **IFG**: inferior frontal gyrus. **HG**: Heschl's gyrus. **PT**: planum temporale. **Hc**: hippocampus. **IC**: inferior colliculus. **MGB**: medial geniculate body.

1.4.2.4 Sequencing and predictability

Nastase et al. (2014) measured BOLD responses to both auditory and visual sequences with varying levels of disorder. Domain-specific findings were stronger than domain-general effects, with auditory disorder encoded by STS and lateral temporal cortex, as well as left IFG; a region often implicated in language-related processes and artificial grammar learning. Auditory sequence encoding may also occur in traditionally motor-related areas, even when behavioural motor responses are not acquired, such as the SMA (Nastase et al. 2014) and basal ganglia (Janata and Grafton 2003; Green et al. 2018).

1.4.2.5 Neuromodulatory mechanisms

Neuromodulators play a crucial role in influencing global brain state, to rapidly deploy neuronal plasticity optimally according to current uncertainty levels. Acetylcholine (ACh) is thought to signal known levels of uncertainty, as estimated from contextual cues. In terms of the brain's hierarchical generative model, ACh specifically communicates the level of uncertainty of top-down predictions, promoting a greater influence of bottom-up input on model updates (as proposed by Dayan and Yu 2002). Neuronally, ACh may enhance the strength of feedforward synaptic connections through increasing gain of superficial pyramidal cells, increasing the influence of new sensory information (Moran et al. 2013). Corroborating this account, direct manipulation of cholinergic activity, by applying ACh to IC neurons, reduces stimulus-specific adaptation (Ayala et al. 2016), in a manner consistent with altering the balance of information flow in favour of bottom-up activity.

The occurrence of unexpected uncertainty, in other words a surprising occurrence which is not accounted for by expected noise in the environment, is associated with a phasic increase in noradrenaline (Payzan-LeNestour et al. 2013). Zhao et al. show that pupil dilation, an indicator of noradrenergic activity, occurs when auditory sequences transition from a more-ordered to a less-ordered state (Zhao et al. 2018). Noradrenaline acts as an 'interrupt' signal, inducing a rapid increase in learning rate (Yu and Dayan 2005).

1.5 Thesis Outline

1.5.1 A more complex model of auditory regularity

This thesis aims to address some of the lesser-studied aspects of learning from regularity in the auditory environment. With so much claimed for the explanatory power of predictive coding, and with a substantial part of the evidence for its operation in the auditory system coming from very simple models of regularity engendered by the oddball paradigm and its variants, it is important to bridge the gap between simple repetition and more intricate forms of regularity. This thesis represents a step towards real-world signals whilst still retaining the distilled purity of tone-pip paradigms.

The stimuli themselves will be presented in more detail in Chapter 2 and in each experimental chapter. The stimuli are adapted from those used by Barascud et al. (2016). Briefly, the stimuli are rapid sequences of tone pips spread across a wide range of pitches. The individual 'notes' in the sequences are arranged according to certain rules, which remain stable throughout each exemplar. Regular stimuli (**REG**) are arranged in cyclically repeating patterns, representing a fully predictable rule, and are contrasted with random (**RAND**) sequences. RAND sequences can also vary in the number of unique elements, representing probabilistic acoustic structure with different levels of uncertainty.

These stimuli are presented at a much faster inter-tone interval than tones in the oddball or streaming paradigms, and are fully adjacent with no silent inter-tone interval. In contrast to statistical learning paradigms, the regularities I use are extractable over the course of a single trial; only a few seconds. There are several advantages of this property. Firstly, we hope to specifically probe automatic processing of regularity rules, distinct from any conscious reasoning of sequential structure. Secondly, the flux of the spectro-temporal information in natural stimuli can be very fast, with changes occurring without an intervening silent gap. Thirdly, it allows us to measure response evoked by the sequence as a whole, which is a direct measure of how auditory context is encoded, which is a complementary approach to the common auditory evoked response paradigms where responses are extracted to individual tones.

1.5.2 Predictability and attention

One explanation proposed for the increased MEG response to regular patterns relative to random ones shown by Barascud et al. (2016), was that regular patterns capture attention to a greater extent. This explanation is compatible with the proposed link between precision and attention in gPC (§1.3.2.3). In Chapter 3, EEG responses to REG and RAND stimuli were recorded, whilst subjects attended to a visual task. These EEG responses demonstrate fast recognition of predictable patterns. As previously seen in MEG, this was revealed by an increase in responses to regular relative to random sequences. One explanation given for this effect was that predictable sequences preferentially capture attention thanks to their higher precision. Behavioural experiments investigate attentional capture by stimulus structure, suggesting that regular sequences are easier to ignore, and do not exogenously capture attention.

1.5.3 Predictability and repetition

Responses to repetitive stimulation generally exhibit suppression. This repetition suppression is thought to form a building block of regularity learning. However, the patterns used in this thesis show the opposite effect, where predictable patterns show a strongly enhanced brain response, compared to frequency-matched random sequences. One possible reason for this discrepancy is that the simple tone pip sequences used in most MMN studies confound predictability with repetition. Furthermore, the differences in temporal properties of the sequence, where MMN studies use an ITI of several hundred milliseconds, but REG and RAND are presented much faster with no intervening silent gaps, may play a role. Chapter 4 presents an EEG study which reconciles auditory sequence predictability and repetition in a single paradigm, by including sequences of tone pips at a single frequency but with the same temporal properties as REG and RAND. Results indicate a system for automatic predictability monitoring which is distinct from, but concurrent with, repetition suppression.

1.5.4 Predictability and deviance detection

The brain's internal model can be investigated via the response to rule violations, as most prominently utilised by the MMN literature. Such studies lend support to the idea that deviations from expected auditory input are encoded in evoked responses, and that the deviance response reflects a precision-weighted prediction error. Chapter 5 and Chapter 6 present behavioural and EEG experiments where violations are inserted in REG and RAND sequences. REG and RAND represent a high-precision and a low-precision context respectively, and as such it was expected that equivalent deviant events would be preferentially processed and lead to a larger response in REG than in RAND sequences. Outlier tones within regular sequences indeed evoked a larger response than matched outliers in random sequences. However, this effect was not present when the violation comprised a silent gap.

1.5.5 Changes in predictability

Chapter 7 concerns the ability of the brain to update an existing model. Regular patterns transitioned to a different rule, keeping the frequency content constant. EEG shows a mismatch response, then a period of adjustment to the rule change, followed by a return to tracking the predictability of the new sequence as reflected by the sustained response magnitude.

Chapter 2. General Methods

2.1 Stimuli

The stimuli used throughout this thesis are sequences of tone-pips arranged according to various rules. The aim of the paradigm is to probe brain mechanisms which track the pattern, rather than simple acoustic properties, of the stimulus. Similar stimuli have been used previously (Chait, Poeppel, de Cheveigné, et al. 2007; Chait, Poeppel, and Simon 2007; Chait et al. 2008; Patel and Chait 2011; Chait et al. 2012; Jaunmahomed and Chait 2012; Barascud et al. 2016; Barczak et al. 2018; Herrmann and Johnsrude 2018).

2.1.1 Tone-pip sequences

Each tone pip is 50-ms in duration, and is ramped on and off with a raised cosine ramp with 5 ms rise and fall time, to minimise transients. Tone pips are concatenated without any gap. The frequencies of the tone pips are drawn from a pool of logarithmically-spaced values (12% frequency increments). The frequency ranges covers between approximately 200 and 3500 Hz, which coincides with important frequencies in speech, and the peak of human auditory thresholds (Suzuki and Takeshima 2004). Sound sequences are between 3 and 4.5 seconds in duration, and are always unique on every trial. This ensures that the processes tapped are related to short-term echoic memory and initial regularity extraction, rather than memory retrieval. Sequences of frequencies are either regularly cycling (REG) or random (RAND), each composed of a number *alph* of frequencies, forming the 'alphabet' used in each sequence. For example REG10 refers to a sequence of 10 frequencies, which repeats every 10 tones. For each trial, a number *alph* of values are drawn from the frequency pool to generate a *sub-pool*. In Chapter 3, the selection was drawn with replacement, as per Barascud et al. (2016). In Chapter 4 and onwards, *alph* frequencies were drawn *without* replacement. The advantage of the latter technique is that a sequence generated from a sub-pool of size *alph* is guaranteed to contain exactly *alph* unique frequencies. With the former technique, the true alphabet size was likely to be lower.

2.1.2 Randomisation

To generate REG, *alph* frequencies are drawn from the pool (with replacement in Chapter 3; without replacement elsewhere). These frequencies are then iterated until the sequence is complete. To generate RAND, two different methods were used. In Chapter 3, to keep the method as similar as possible to previous work, each frequency was drawn from the sub-pool at random with replacement. However, for later chapters, the randomisation procedure was modified to ensure both REG and RAND contained the same number of tones at each frequency over the entire stimulus duration. Each RAND sequence was obtained by simply generating a REG first, then shuffling the order. A further control was applied in Chapter 4 onwards. REG, being a repeated cycle of unique tones, never contains adjacent tone pips at the same frequency. RAND, on the other hand, is expected to contain adjacent repeats on 1/*alph* of the sequence elements. This could lead to differential adaptation in REG and RAND, as the repeated tone is known to evoke a smaller response (see §1.2). To avoid this, RAND sequences were generated by shuffling REG sequences up to 1000 times and the sequences with fewest adjacent repeats were chosen. This proves too strict a set of constraints for smaller *alph*, so this control was only used for alphabet size of 10 or more in Chapter 4 onwards.

2.1.3 Ideal observer modelling of stimuli

Stimulus sequences used throughout the thesis were analysed to generate an estimate of how surprising each new tone is, given a learnt model of the statistical structure over the previous tones presented in the experiment. The motivation is to compare brain responses to an 'ideal' observer. This approach has previously been used for REG and RAND stimuli, and the model output closely replicates the dynamics of the brain response measured with MEG (Barascud et al. 2016). The sequences were analysed using an unsupervised learning model of musical expectation: Information Dynamics Of Music (IDyOM. Pearce 2005; Pearce et al. 2010). The model optimally forms statistical predictions about the likely continuation of a sequence of notes given a memory of which transitions between adjacent tones have occurred previously.

IDyOM was designed for modelling musical melodies, but is also suitable for use with non-musical stimuli such as these. This is because the modelled data feature is simply a sequence of note identities (expressed in an arbitrary unit such as musical notation, or musical intervals, or here as frequencies in Hz); timing information and other score markings are not included in generating predictions. IDyOM models the sequence as a variable-order Markov chain, where the probability of a note at a given timepoint is conditioned on the identity of the previous *n* notes, where *n* is the order of the Markov chain. The model incorporates predictions from multiple different simultaneously estimated Markov chains at different values of *n*, allowing it to learn higher-order dependencies requiring a long stimulus history to predict, as well as local dependencies.

The model was presented with the sequences in a stimulus set one note at a time, paralleling the information available to a real subject. The present thesis uses a particular configuration of the model, LTM+, which begins with an empty model, then progressively learns over the entire corpus of sequences presented thus far in the simulated experimental session, updating the model after each note. This configuration was found by Barascud et al. (2016) to most closely correspond with brain responses.

The model infers multiple variables, but for the purposes of this thesis, the information content associated with each tone was used as a measure of surprise. At each note x_t in the melody, occurring at position t in the sequence, information content IC is defined as the negative log-probability of x_t given the preceding sequence context under the model M; giving the expression $IC = -\log(P(x_t|x_{t-n}, M_t)$

This is equivalent to surprisal under the internal model fit by the algorithm. Given that EEG evoked responses are thought to reflect prediction error (Friston 2005), it is hypothesised that EEG response power will closely track this information measure. Indeed, evoked response magnitude and behavioural reports of surprise correspond closely to information content from this model (Egermann et al. 2013; Omigie et al. 2013; Agres et al. 2017).

2.2 Presentation Conditions and Equipment

For behavioural experiments, sounds were presented through circumaural headphones (Sennheiser HD595) using a soundcard (Roland Tri-Capture) attached to a Windows laptop computer. For EEG experiments, participants listened to stimuli through non-metallic extended earphones (3M E-A-Rtone) such that there is minimal magnetic field interference from the speaker driver with the EEG recording. All experiments took place in a sound-proof booth, at a listening level chosen to be comfortable for each participant. Rests were provided at most every ten minutes. Auditory stimuli were binaurally presented, and controlled by the Psychophysics Toolbox extension in Matlab (Kleiner et al. 2007).

EEG recordings were made using a Biosemi system (Biosemi Active Two AD-box ADC-17, Biosemi, Netherlands) with either 64 or 128 Ag-AgCl electrodes. The recording was obtained at 2048 Hz and subsequently downsampled.

All subjects were aged between 18 and 35 years, fluent English speakers with normal or corrected-to-normal vision. Participants were reimbursed for their time. None reported a history of hearing impairment or neurological disorder. All experimental procedures reported in this thesis were approved by the research ethics committee of University College London, and written informed consent was obtained from each participant.

2.2.1 Visual decoy tasks

A visual decoy task was used for the EEG experiments in Chapters 3 and 7. The visual task was displayed on a separate computer, using the Cogent 2000 toolbox for MATLAB (www.vislab.ucl.ac.uk/cogent.php). The timing was not correlated with that of the auditory stimuli. For each trial, 3 colour photographs of landscapes were shown for 5 s each, and images faded gradually from one image to the next to minimise visual transients, which could themselves produce an evoked response. Subjects were instructed to press a keyboard button if the first and third image within a trial were identical (10% of trials), and to withhold a response otherwise. There were a total of 288 unique images. The inter-trial interval was jittered between 2 and 5 s. The session was split into four consecutive blocks. Feedback (number of hits, misses and false alarms) for the visual task was provided at the end of each block. Behavioural performance on the visual task was at ceiling.

For Chapters 4, 5 and 6, no visual task was used. Instead, subjects watched a film of their choice with subtitles and the volume muted. Subjects were instructed to relax and minimise movement, particularly jaw-clenching and facial movements.

2.3 EEG

Brain responses throughout this thesis were measured with electroencephalography (EEG). This technique measures electrical activity on the scalp with sub-millisecond precision, which is advantageous for studying the rapid processes of sequence extraction from sound. As reviewed in Chapter 1, there is a rich history of using EEG to learn about how the brain processes sound in particular, and patterns in general.

2.3.1 Physical principles

The link between neuronal activity and measured EEG is complex. EEG simply measures electric potential differences between sensors on the scalp. Neuronal activity influences the electric potential in each neuron’s surroundings in a myriad of ways: transmembrane currents, action potentials and postsynaptic currents, though most of the former cancel out locally. Electrical currents conduct differently through the various tissues of the brain, skull and scalp, leading to idiosyncratic contributions of each source of activity over multiple sensors, as well as simply dropping in power with the inverse square of distance travelled. As a result, each sensor on the scalp measures the net effect of any number of neurons’ activity, as well as other sources of electrical activity in the environment. For a neuronally-generated current to reach the scalp with measurable amplitude (approximately in the order of 0.1 µV), many neurons must be simultaneously active, with an estimated lower bound of some 10-50,000 neurons (Murakami and Okada 2006). These neurons must also be approximately aligned spatially: axons facing in different directions will each generate electrical current locally, but opposing directions will cancel out on the macroscopic scale (Ahlfors et al. 2009). The arrangement of pyramidal neurons in the cortex is highly regular, with a predominantly perpendicular arrangement with respect to the local cortical surface, so such favourable summation does thankfully occur.

2.3.2 Source modelling

The primary generators of neural current, even though they have a spatial extent, can be adequately modelled as a point source with an oriented net current: a dipole. An infinite number of such dipoles (characterised by their location, orientation and power) can explain a given pattern of neural activity. To (approximately) solve this ill-posed inverse problem requires constraints from assumptions grounded on other sources of knowledge.

Detailed information can be gleaned from a structural MRI on an individual subject’s head shape, and the thickness and shape of individual tissues, and the configuration of their cortical surface. Such information allows the estimation of a realistic *forward model*, namely a mapping from sources in the brain to potential measured at each sensor.

Various approaches for estimating an inverse solution are available which differ in the underlying assumptions (reviewed by Yao and Dewald 2005; Grech et al. 2008; López et al. 2014). In this thesis, the source inversion uses Minimum Norm Estimation (MNE; Dale et al. 2000) as implemented in the Fieldtrip toolbox (www.fieldtriptoolbox.org/; Oostenveld et al. 2010). This method simultaneously fits multiple dipole strengths over the whole brain, allowing any combination to be simultaneously active. The solution is constrained by selecting the combination of dipoles with the lowest total energy which adequately explain the sensor-level data. The distributed nature of the solution is particularly ideal for locating the differences between REG and RAND; as from previous work (Barascud et al. 2016) it is expected that the response to these sequences involves multiple sources.

2.3.3 Noise

The neural electrical potentials measured by EEG are tiny, and can be dwarfed by other sources of electrical noise (reviewed in Luck 2005). For instance, the movement of facial muscles and eye movements both produce artefacts many times larger than fluctuations in the EEG. Mains electricity produces a 50Hz signal. Slow drifts in the EEG signal are known to occur due to changes at the skin-electrode interface during the recording (Huigen et al. 2002), and these drifts are generally uncorrelated across channels, precluding the use of component analysis techniques to remove them (de Cheveigné and Arzounian 2018). The impact of noise is mitigated by event-related averaging (Luck 2005), as most noise sources are assumed to be uncorrelated with the cognitive process of interest. In this thesis, around 100 repetitions of each condition were recorded for each subject. Further signal-to-noise improvement was enabled by the use of various preprocessing techniques.

2.3.4 Preprocessing

Data were analysed using the Fieldtrip toolbox (www.fieldtriptoolbox.org/; Oostenveld et al. 2010) for Matlab (2015a, MathWorks). The analysis pipeline described below was developed throughout the course of the PhD research. Therefore, earlier studies do not make use of all the preprocessing techniques. The true chronological order of the chapters was 3, 5, 6, 4 then 7; however for flow of the thesis argument these were re-ordered.

2.3.4.1 Filtering

Most EEG preprocessing pipelines include a high-pass filter to remove the slow drift in the signal and improve the signal-to-noise ratio (Huigen et al. 2002; Kappenman and Luck 2010). Commonly used cut-off values for auditory evoked responses are 1 or 2 Hz. However, the slow dynamics of the sequence-evoked response are of particular interest here as they may represent the tracking of predictability, such as has been reported with similar stimuli previously (Barascud et al. 2016; Southwell et al. 2017). Also, the introduction of filter-related distortion of evoked responses becomes a concern with increasingly high cut-off frequencies (e.g. above 0.3Hz; Tanner et al. 2015). A much lower cut-off of 0.1 Hz is used here to preserve slow dynamics whilst mitigating some of the slow drift. However, for comparison to 'traditional' evoked responses such as the N1 and MMN, additional analyses of transient evoked potentials here use a high-pass cut-off of 2Hz. Unless otherwise specified, high-pass filtering was applied on the continuous dataset, before splitting the data into epochs, in order to minimise the impact of filter edge artefacts (de Cheveigné and Arzounian 2018). Unless otherwise specified, a fifth-order Butterworth filter was applied in forward and reverse directions, implementing a zero-phase-shift (acausal) filter. The acausal filter is the default method in Fieldtrip, and has the advantage that it does not delay the apparent latency of peaks, however as its output at a given sample depends on both the past and the future, it can also lead to the appearance of effects beginning at an earlier latency than in reality due to influence from future peaks (de Cheveigné and Nelken 2019).

Data were also low-pass filtered prior to down-sampling, at or below the Nyquist frequency, to avoid aliasing artefacts. The focus of this thesis is relatively slow EEG components, so a 30 Hz low-pass filter is applied prior to statistical analysis. This also removed 50 Hz noise from electrical mains and high-frequency muscle artefacts.

2.3.4.2 Detrending

Signal-to-noise ratio can be improved by using a high-pass filter; however such filtering may introduce distortions. For instance, discontinuities or spikes in the raw data can introduce ringing artefacts (de Cheveigné and Arzounian 2018). Filtering can also affect the apparent timing of response peaks (see de Cheveigné and Nelken 2019), and, crucially for the sustained response effects, filters may distort true slow components in the EEG signal (Tanner et al. 2015; Widmann et al. 2015; Maess et al. 2016). Detrending is an alternative option to preserve these slow dynamics whilst removing the electrode drift. Here, a polynomial function of arbitrarily-specified order is fitted to the data in each channel, then subtracted. However, this is also vulnerable to introducing distortions if there are discontinuities or glitches in the data. *Robust* detrending is a variant of this process, where a weighting is applied to each sample in the timeseries, to determine its influence on the polynomial fitting (de Cheveigné and Arzounian 2018). Weighting can be used in several ways to render the process more resilient to artefacts. Firstly, a thresholding can be applied such that all samples exceeding a set number of standard deviations from the mean are given a weight of 0, and 1 otherwise. In this thesis, a threshold of 3 standard deviations was used. Secondly, periods of the signal known to contain no brain responses of interest, such as the inter-trial interval, can be uniquely used to anchor the fit, applying a low or zero weighting to the peristimulus interval. This helps prevent the fitting of, and removal of, brain evoked slow potentials. Thirdly, known glitches in the data can be down-weighted (if not already addressed by the thresholding).

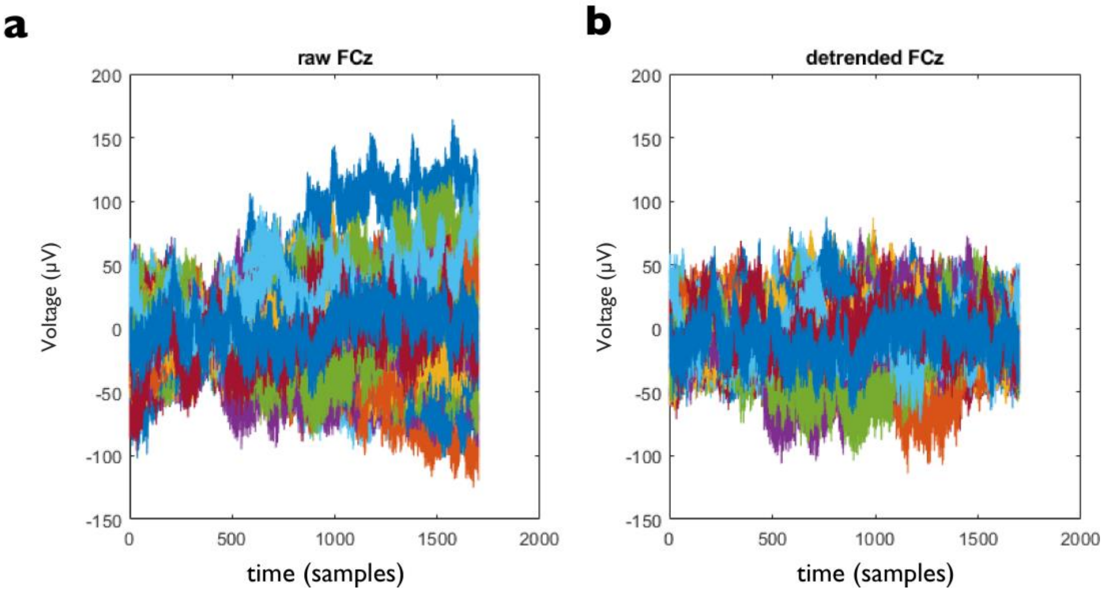

Figure 2.1 Example raw and detrended data
Data from a single block recorded from a subject taking part in the EEG experiment in Chapter 7. Each coloured line is the signal at channel FCz over an entire epoch. **a:** raw data. **b:** data after detrending.

2.3.4.3 Artefact rejection

Remaining artefacts can be removed by eliminating entire epochs and channels from the dataset. Outlier channels and trials were removed manually with the aid of Fieldtrip's visual artefact rejection tool, which displays summary statistics for each trial and channel: variance, range, kurtosis, z-score and maximum absolute values. The user then manually selects individual trials and channels to remove on this basis. The proportion of trials retained according to these criteria varied substantially between experiments, probably reflecting different levels of noise in the datasets and individual differences in signal quality, but was never over 25%. As the number of trials generally outweighed the number of channels, first trials were rejected then channels.

In the final stages of preprocessing, just before computing the average over trials, data from rejected channels were then reconstructed as the average of the signal from all immediately neighbouring channels. Subsequently, data were re-referenced to the average over all channels and baseline-corrected by removing the mean voltage during the interval prior to the event of interest.

2.3.4.4 Independent component analysis

Independent component analysis (ICA) can be used to separate sources of activity which are statistically independent from one another, including sources of noise and independent neural signals. ICA decomposes the data into a linear combination of component timeseries which share minimum mutual information. Here, it was used to identify eye-blink, eye-movement and (occasionally) cardiac artefacts. The `runica` algorithm implemented in EEGLAB, included in Fieldtrip, was used (Makieg et al. 1997; Delorme and Makeig 2004). Artefact components were visually identified from their topographies and timeseries and the relevant components omitted when projecting back into the sensor domain.

2.3.5 Denoising source separation

Further signal-to-noise improvement was achieved using *denoising source separation* (DSS; Särelä and Valpola 2005; de Cheveigné and Simon 2008; de Cheveigné and Parra 2014). This allows separation of the stimulus-evoked response from various sources of noise. DSS finds a linear decomposition of the dataset into a set of mutually-orthogonal components. Each component is described by a ‘spatial filter’: a time-invariant weighting over sensors, and is associated with a bias score, quantifying how much in common the component has with a bias filter, which determines what is considered ‘signal’ to be emphasised. The component selection represents a rotation of the original dataset to maximise variance in the components aligned with the bias filter. There is freedom to choose any bias filter, for example to minimise power line noise, maximise narrowband oscillatory activity or maximise evoked power (de Cheveigné and Parra 2014). The final stage is to select the best *k* components, as ordered by their bias score, to project back into the same dimensionality as the original dataset, yielding ‘clean’ data with an enhanced SNR.

DSS uses two steps of principle component analysis (PCA), but unlike PCA it orders components by similarity to the bias function rather than simply proportion of response power explained, which when used as a denoising technique assumes the signal of interest is more powerful than the noise. Unlike ICA, DSS produces an ordered set of components according to the criterion set out in the bias function.

Here, DSS was applied to maximise the reproducibility of scalp activity patterns across trials, for each subject. The bias filter is simply the trial-averaged data for that subject. In order to avoid circularity in the statistical inferences that were eventually drawn from the data, DSS was applied to trials from all conditions pooled together, thereby ‘blindly’ improving the SNR for all conditions together.

DSS to maximise repeatability is only able to detect exactly time-locked repeatable features, and as such may be ineffective for later, less stereotyped brain responses. DSS will also thus boost the contribution of any repeatable noise. One such source of noise would be any slow drift not previously accounted for: the slow drifts tend to be fairly consistent across trials so show up in the event related average used to derive the bias function. However, as shown in the example plot below, it can have a strong impact on the data quality.

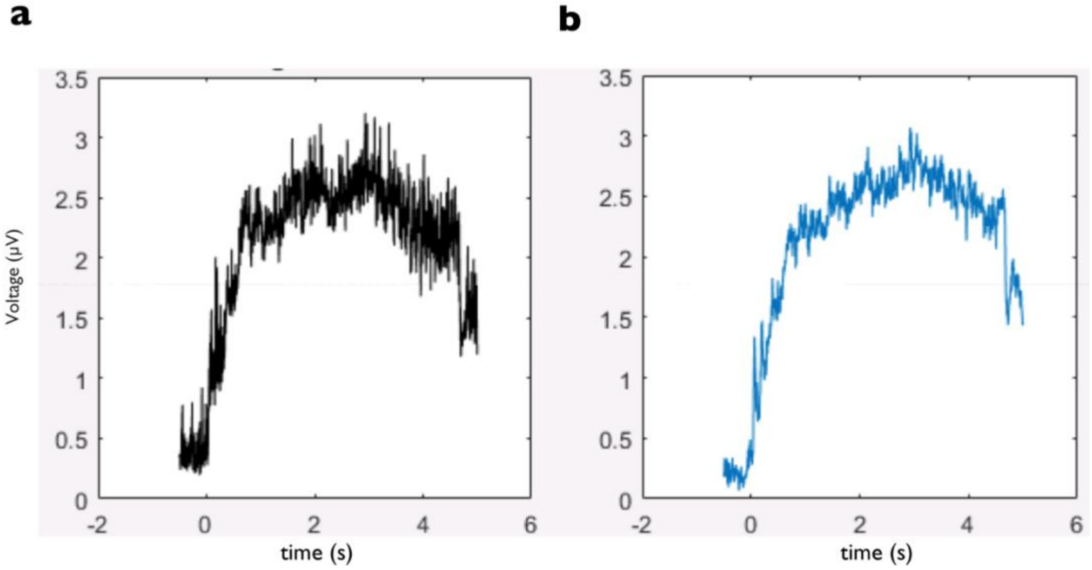

Figure 2.2 DSS: before and after

Trial-averaged data **a:** before and **b:** after applying DSS to maximise evoked repeatability. Data from a single subject taking part in the EEG experiment in Chapter 7. Here, the bias filter was the average evoked response power during the first 500 ms after stimulus onset.

2.3.6 Quantifying evoked responses

The principle measure of the evoked responses used in this thesis is the global field power, calculated as the root-mean-square (RMS) of the electrode readings at each time point over all channels (Murray et al. 2008). This quantifies the response strength over the whole scalp, and retains the same units (μV) as the responses in individual channels. The RMS is used for analyses examining the sustained response magnitude. In previous work using MEG (Barascud et al. 2016) and the EEG responses in this thesis (see Figure 2.3), the sustained portion of the sequence-evoked response, from around 300 ms, shows no zero-crossings. All channels show a similar shaped response. The RMS captures the overall dynamics of the response, substantially reducing the dimensionality, without disregarding any of the channels. Group averages of the RMS over all subjects are also computed as the root-mean-square rather than the arithmetic mean (Murray et al. 2008).

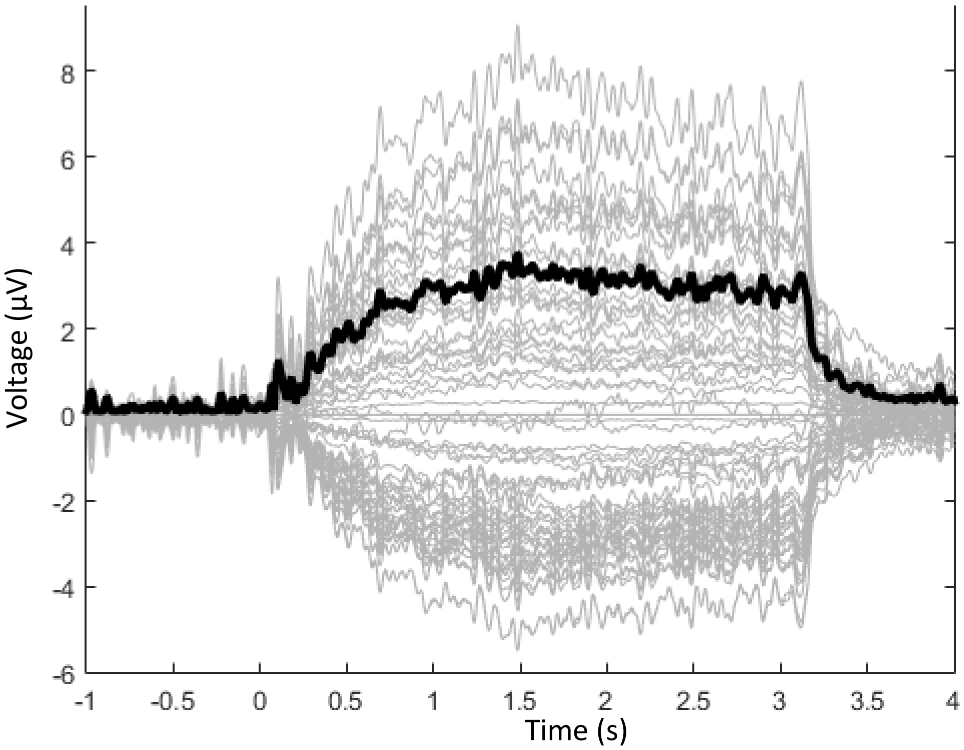

Figure 2.3 EEG butterfly and RMS

Example plot of the sustained response to a REG5 sequence (taken from Chapter 3). Individual channels are shown in grey, and the root-mean-square average over all channels is shown in black.

2.4 Statistics

2.4.1 Multiple comparisons

EEG data is high-dimensional, yet the theoretical conclusions ultimately drawn from it are low-dimensional. Qualitative conclusions drawn from the high-dimensional dataset, such as "condition A shows decreased evoked response power as compared to condition B", are based on a multitude of statistical tests performed at each point in space and time. When multiple statistical tests are carried out, the chance of obtaining a false positive on one of the tests increases; this is known as the *multiple comparisons problem* (MCP). If a conclusion is drawn from the outcome of such a *family* of statistical tests, each with threshold α for significance of $p = 0.05$, solving the MCP amounts to controlling the overall *family-wise error rate* (FWER) under the null hypothesis to be no greater than α.

There are different approaches to this, which fall on a spectrum of more conservative, and therefore more likely to falsely show no effect, to more lenient, and therefore more likely to falsely show an effect. Bonferroni correction is the most conservative, simply dividing the critical α by the number of statistical tests undertaken. However, if there is some co-dependence in the data, such as correlation over space or time, this is overly conservative. Both parametric and non-parametric methods exist to solve the MCP whilst taking account of this structure in the data to perform a less conservative correction.

2.4.2 The cluster-based permutation test statistic

This thesis uses a non-parametric procedure to control the FWER, taking account of the co-dependency between nearby channels and time samples in the data. The advantage of using a non-parametric approach is that no assumptions are made regarding the underlying statistical distribution describing the dataset. This procedure is described in (Maris and Oostenveld 2007). For example, to compare two conditions A and B, the difference at each sample is quantified by some metric, such as the t-statistic. Then a threshold is applied, identifying samples which show a difference greater than the threshold value. Contiguous values exceeding the threshold are subsumed within a single cluster. Then, correction for multiple comparisons is performed over these clusters, which are less numerous than the original samples. A cluster-wise statistic is computed; commonly the sum or maximum of the individual metrics. A null distribution of the value of the cluster-wise statistic is estimated using a permutation approach: condition labels are randomly permuted before repeating the process of computing the cluster-wise statistic. This is repeated for some large number of iterations, and from this a probability of obtaining a test statistic greater than the one obtained from the true dataset is computed. If this probability is smaller than the desired α, there is a significant difference between conditions, controlled for a family-wise error rate of 0.05.

Chapter 3. Predictability and Attention

3.1 Summary

In this series of behavioural and electroencephalography (EEG) experiments we investigate the extent to which repeating patterns of sounds capture attention. Work in the visual domain has revealed attentional capture by statistically predictable stimuli, consistent with predictive coding accounts which suggest that attention is drawn to sensory regularities. Here, stimuli comprised rapid sequences of tone pips, arranged in regular (REG) or random (RAND) patterns. EEG data demonstrate that the brain rapidly recognises predictable patterns manifested as a rapid increase in responses to REG relative to RAND sequences. This increase is reminiscent of the increase in gain on neural responses to attended stimuli often seen in the neuroimaging literature, and thus consistent with the hypothesis that predictable sequences draw attention. To study potential attentional capture by auditory regularities, we used REG and RAND sequences in two different behavioural tasks designed to reveal effects of attentional capture by regularity. Overall, the pattern of results suggests that regularity does not capture attention.

3.2 Statement of Contribution

This chapter has been adapted from a published paper: **Southwell R, Baumann A, Gal C, Barascud N, Friston K, Chait M. 2017. Is predictability salient? A study of attentional capture by auditory patterns. Philos Trans R Soc Lond, B, Biol Sci. 372:20160105–20160111.** For the purposes of this thesis, I added sections to show modelling of sequence information content using IDyOM (§3.4.1, Figure 3.2), and additional discussion of results with more recent references (Addendum, §3.8).

AB, CG, NB, MC and I designed the experiments. CG and I acquired the data for Experiment 1, and I analysed the data. I acquired and analysed the data for Experiment 2. AB acquired and analysed the data for experiment 3. AB, CG, NB, KF, MC and I interpreted the results. MC and I wrote the article.

3.3 Introduction

The human brain is highly sensitive to patterns in sensory input (Saffran et al. 1999; Chait, Poeppel, de Cheveigné, et al. 2007; Turk-Browne et al. 2009; Barakat et al. 2013; Wang et al. 2015). A growing body of work in vision (Fiser and Aslin 2001; Turk-Browne et al. 2009), touch (Conway and Christiansen 2005b), language (Huigen et al. 2002), and audition (Gebhart et al. 2009; Agus et al. 2010; Bendixen, Schröger, et al. 2012; Garrido et al. 2013; Paavilainen 2013; Bendixen 2014) has demonstrated that subjects rapidly and automatically learn complex sensory statistics, and that these are exploited to improve perceptual inference, even when outside of conscious awareness. This capacity is often interpreted as a fundamental element of the predictive mechanisms which are proposed to constitute the principal substrate of perception (Friston 2005; Clark 2013; Hohwy 2013).

In hearing, automatic sequence learning has commonly been studied via the mismatch negativity (MMN); an electrophysiological marker for the processing of sounds that break an established rule (Paavilainen 2013). MMN to sequence violations has provided (indirect) evidence that the auditory system can learn complex rules governing sequences (Nordby et al. 1988; Bendixen 2014). The repetition positivity, which increases with the number of repeated stimuli, is another neural marker of simple regularity extraction (Baldeweg 2006). Recently, Barascud et al. (2016) provided direct evidence of the process of regularity extraction in more complex tone sequences. They used rapid sequences of tones with frequencies changing in a regular, cyclical pattern, and matched sequences of tones arranged in a random order. Behavioural reaction times and neural response dynamics indicated rapid recognition of regularity, on par with the latency predicted from an ideal observer model.

Learned knowledge about regularities, whether from low-level statistical learning or conceptual understanding of the phenomena causing sounds, enables predictions to be formed about future sensory input (Winkler et al. 2009). Such expectations improve behavioural performance in predictable contexts; for example, by orienting resources to a point in time when a stimulus is expected (Nobre et al. 2007), or by facilitating selective attention and segregation of concurrent sound streams (Denham and Winkler 2006; Bendixen et al. 2010; Andreou et al. 2011; Bendixen 2014). In addition, recognition of regularities can aid detection of changes in the environment, which causes sensory input that is in disagreement with these predictions (Bendixen, Schröger, et al. 2012; Schröger et al. 2013; Barascud et al. 2014). It has been proposed that the same predictive mechanisms underlie both the detection of regularity violations and auditory scene analysis (Winkler et al. 2009; Schröger et al. 2013).

3.3.1.1 Attention

Attention allows the prioritisation of useful streams of information for further processing. Within this context, the relationship between predictability and attention is attracting increasing research interest (Feldman and Friston 2010; Kok, Rahnev, et al. 2012; Chennu et al. 2013; Summerfield and de Lange 2014; Summerfield and Egner 2016). However, there is disagreement as to whether it is unpredictable, surprising events (Pearce and Hall 1980; Itti and Baldi 2009) or predictable ones (Mackintosh 1975) that are the most informative in scene analysis, and should therefore (in the parlance of Itti & Kock Itti and Koch 2000; 2001) be flagged as more 'salient' and attract selective attention (Dayan et al. 2000; see also Pearce and Mackintosh 2010). In visual studies, it has been shown that learning of regularities helps guide attention to expected locations (Chun and Jiang 1999; Summerfield et al. 2006) and features (Chun and Jiang 1999; Chalk et al. 2010). Zhao et al. (2013) recently proposed a framework for attentional guidance whereby automatically-learned regularities in the sensorium bias attention, even if not relevant for performing a task, and demonstrated this to operate in guiding visual search. They presented sequences of abstract shapes; the order of which was statistically structured at a particular location in the search array and random at others. This was followed by a static visual search array. Reaction times were faster to targets presented at the statistically structured array location, despite the regularity carrying no predictive information as to the target location.

Zhao et al. (2013) suggest that the prioritisation of regular features is a means to focus resources on stable aspects of the world, which can then be learnt. The notion that the brain is 'hardwired' to prioritise regularities is at the heart of popular models of the brain as a statistical organ of prediction. The expected precision of bottom-up information streams plays a vital role in such predictive processing accounts (Friston 2005; Feldman and Friston 2010). Reliable prediction errors are up-weighted in proportion to their expected precision, thereby refining the brain's generative model based on the most informative streams (Feldman and Friston 2010).

3.3.1.2 Increased responses to regular stimuli: exogenous attention?

These ideas may help explain an intriguing recent finding concerning the passive brain response to acoustic patterns. Barascud et al. (2016) found a substantial increase in the neural response to regularly-repeating sound sequences over similar random sequences. This finding seems contrary to a large body of work showing reduced responses to predictable stimuli (Haenschel et al. 2005; Alink et al. 2010; Costa-Faidella et al. 2011; Meyer and Olson 2011; Todorovic et al. 2011; Todorovic and de Lange 2012). The proposed explanation for the discrepancy is that, unlike many of the signals used in previous work which often consist of oddball or roving sequences (Haenschel et al. 2005; Baldeweg 2008), the stimuli used by Barascud et al. (2016) were complex auditory patterns where the predictability of sound sequences was not confounded with neural adaptation resulting from repetition of identical sounds. There are several possible explanations of the increased response to regularity. One is that it reflects the engagement of neural circuits for sequence learning, whose activity in addition to the basic response to the stimulus in auditory cortex results in a net increase in magnetic field strength. Another is that the same neural population is simply more active, with the effect resulting from an increased gain on the activity of auditory neurons responding to the stimuli, potentially signalling greater expected precision. At the cognitive level, the result could potentially indicate that subjects were having their attention spontaneously biased towards the regular sounds, even though they were engaging in an unrelated visual task. Indeed, it has been shown that neural response magnitude is enhanced to attended, predictable stimuli in audition (Hsu et al. 2014; Auksztulewicz and Friston 2015), and in vision (Kok, Rahnev, et al. 2012).

In the experiments presented in this paper, we investigate whether regularity captures (exogenous) attention. We use the same stimuli as Barascud et al. (2016), consisting of tone-pip sequences whose frequency pattern is either regularly repeating (REG) or random (RAND; Figure 3.1). In Experiment 1 we demonstrate that the increased brain response to REG relative to RAND also occurs in electroencephalography (EEG). In a series of behavioural experiments, we then investigate the capacity of REG and RAND to exogenously capture attention when they act as auditory distractors (Experiment 2) and test whether auditory regularity biases attention in scenarios where multiple sound streams are attended and task-relevant (Experiment 3). In both of these paradigms, we find no evidence for attentional capture by acoustic regularity.

3.4 Experiment 1 (EEG)

3.4.1 Methods

3.4.1.1 Stimuli

Stimuli (Figure 3.1) were 3000-ms long sequences of 50-ms tone-pips (60 tone pips altogether; each ramped on and off with a 5-ms raised cosine ramp). Tone frequencies were drawn from a pool of 20 logarithmically-spaced values between 222-2000 Hz. A unique sequence was presented on each trial. Sequences were defined by two parameters: *alph* (alphabet size) - the number of frequencies chosen (at random, with replacement) from the pool, and regularity (REG or RAND). In regular (REG) sequences, a sub-pool of *alph* frequencies were chosen from the full pool, and arranged in repeating cycles of length *alph*. Random (RAND) sequences were generated by drawing each tone at random from the sub-pool of *alph* frequencies. REG and RAND sequences of the same *alph* were generated in pairs, using the same sub-pool, such that conditions were matched for the occurrence of each frequency (Figure 3.1). REG conditions used *alph* = 5, 10 and 15; RAND included an additional condition of *alph* = 20 (using the whole frequency pool), yielding 7 conditions (REG5, REG10, REG15, RAND5, RAND10, RAND15 and RAND20). These sequences are too rapid to allow deliberate reasoning of the order of individual tones; nevertheless the repetitions in REG sequences lead to a strong, 'pop-out' percept of a pattern (Barascud et al. 2016). Examples of the stimuli used are provided as Supplementary Materials in the paper (Southwell et al. 2017).

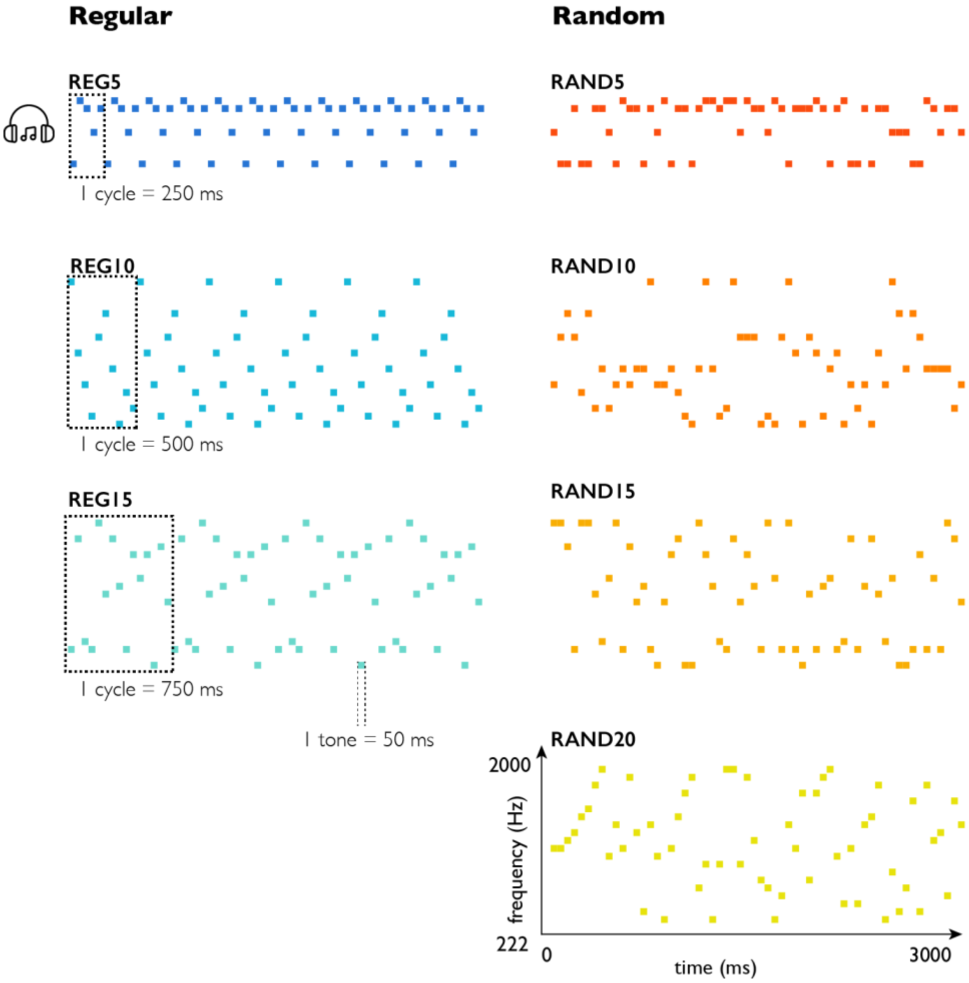

Figure 3.1 Stimuli

Example schematics of the REG and RAND stimuli used. RAND20 contains all 20 frequencies from the pool, in random order. REG5, REG10 and REG15 (left) consist of a regularly repeating pattern of 5, 10 or 15 tones, with frequencies chosen at random, with replacement, from the pool. For each REG stimulus, a matching RAND stimulus, consisting of the same frequencies but in random order, was generated. All stimuli were unique (never repeated) and generated anew for each participant.

3.4.1.2 Ideal observer modelling procedure

The degree of surprise associated with each tone in the sequence reflects the degree to which the sequence has been acquired by the brain. This process was modelled using an ideal observer model of music expectation, IDyOM (Pearce 2005; Pearce et al. 2010; Barascud et al. 2016, see §2.1.3 for more details). The field strength of MEG responses to REG and RAND was shown to closely follow this measure of information content (Barascud et al. 2016), with a negative relationship such that the lower the information content of the tone, the greater the field strength. Here, in a similar approach to that used by Barascud et al. (2016), the stimuli from an entire experimental session, arranged in the same random order, were analysed to estimate transition probabilities between adjacent notes, from which the information content of each tone was derived. The LTM+ model configuration was used, which is initialised with flat prior expectations, but updates the observer model at each successive tone over the entire simulated 'session'. At each tone, the model outputs the probability distribution over the possible frequencies of the next tone, given the current model configuration. From this, the information content of the actual next tone is stored.

3.4.1.3 Ideal observer results

The mean and standard error of the information content of each tone, for each of the sixty positions within the sequence, was estimated with bootstrap resampling with 1000 iterations. Pairs of conditions were compared statistically using a FWER-correction procedure with cluster-wise correction (Maris and Oostenveld 2007) at an alpha level of $p < .05$. Figure 3.2 shows the mean information content, interpretable as the level of 'surprise' elicited by each tone, as a function of sequence position. RAND20 is shown in both plots, representing the stimulus with the most uncertainty, and is used as a baseline against which to asses 'learning' of sequence structure by the model. Indeed, the model shows high information content for RAND20 throughout the sequence, remaining at a similar level of around 5 bits per tone. However, the information content for REG conditions drops rapidly during the first repetition cycle (Figure 3.2a). This drop in information content can be interpreted as the model successfully recognising the regularity in the sequence. For REG5, the information content first significantly differed from RAND20 at 8 tones; for REG10 at 13 tones, and for REG15 at 18 tones. The information content diverges from RAND20 at 3 tones into the first repeating cycle, for all values of *alph.* Following this drop, the information content of REG levels off to between 1-2 bits per tone. The information content for REG5 remains significantly lower than for REG10 and REG15 throughout the sequence, beginning at the eighth tone .

Figure 3.2b shows the information content for the four RAND conditions. This remains at a much higher level throughout the sequence, between 4 and 5 bits per tone. It appears that the information content tends to be smaller for smaller *alph*. This makes sense, as the smaller the alphabet used to generate a sequence, the fewer total possibilities there are for the identity of the next tone, even if the exact frequency is chosen at random. RAND5 has a significantly lower information content than RAND20 throughout the sequence (from tone 3), though RAND10 and RAND15 only sporadically show a significant difference from RAND20.

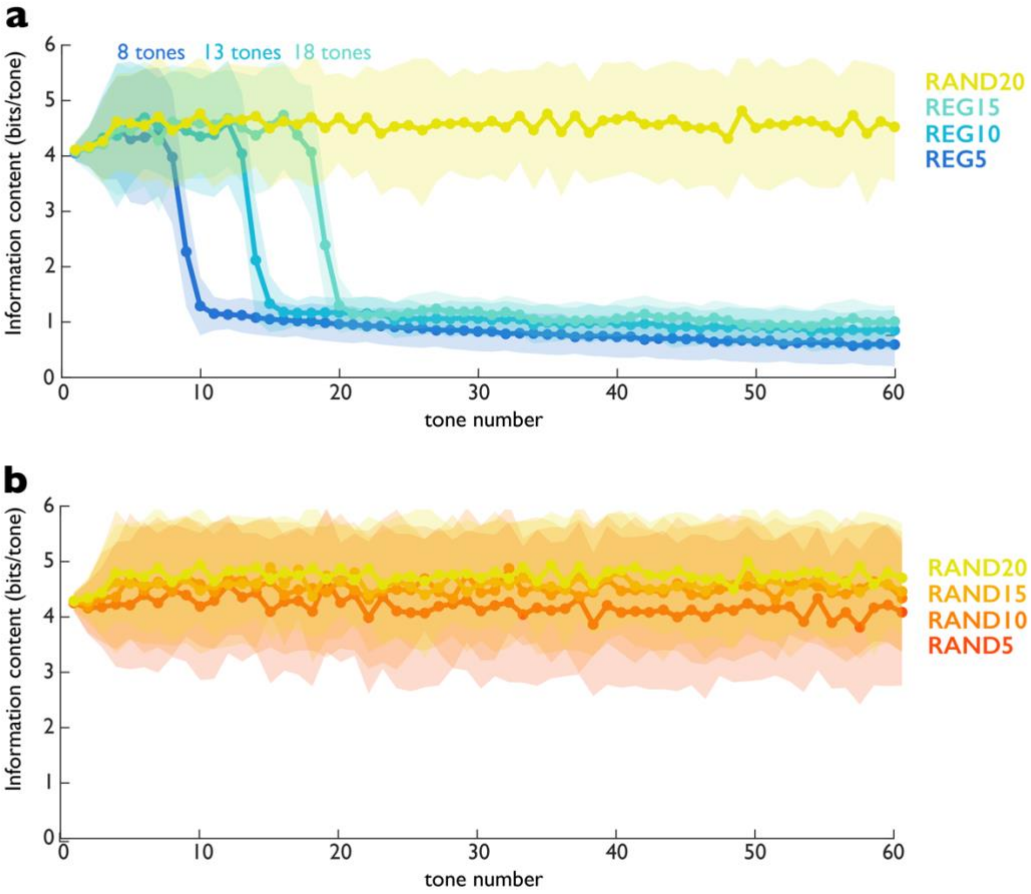

Figure 3.2 Modelled information content

Mean information content per tone, derived from an ideal observer model of frequency expectation, computed over an entire experimental session of 700 trials. Shaded regions represent standard deviation over trials. **a:** REG conditions as compared to RAND20. The information content drops during the first repetition of the sub-sequence, diverging from RAND20 three tones into the repetition. Henceforth, all three REG conditions show low information content per tone, but this is slightly lower for smaller values of alph. **b:** All four RAND conditions show similar information content, remaining roughly constant throughout the sequence. However, as alph decreases, the information content per tone also decreases slightly, representing the predictability arising from the reduced frequency pool.

3.4.1.4 Procedure

The procedure was similar to the MEG experiment described in Barascud et al. (2016). Subjects were engaged in an incidental visual task and were naïve about the nature of the auditory stimuli. Auditory stimuli were presented binaurally with the Psychophysics Toolbox extension in Matlab (Kleiner et al. 2007). In total, subjects heard 700 unique stimuli (100 for each condition). The inter-stimulus interval (ISI) was jittered between 1100 and 1500 ms. The visual task was displayed on a separate computer using Cogent 2000 in MATLAB (www.vislab.ucl.ac.uk/cogent.php). The timing was not correlated with that of the auditory stimuli. For each trial, three colour photographs of landscapes were shown for 5 s each, and images faded gradually from one image to the next to minimise visual transients. Subjects were instructed to press a keyboard button if the first and third image within a trial were identical (10% of trials), and to withhold a response otherwise. Inter-trial interval was jittered between 2 and 5 s. The session was split into four consecutive blocks. Feedback (number of hits, misses and false alarms) for the visual task was provided at the end of each block.

3.4.1.5 Recording & data preprocessing

EEG signals were recorded using a Biosemi system (Biosemi Active Two AD-box ADC-17, Biosemi, Netherlands) with 64 electrodes; at a sampling rate of 2048 Hz. Recording was re-started at each block. Data were analysed with SPM12 (Statistical Parametric Mapping; http://www.fil.ion.ucl.ac.uk/spm/) and Fieldtrip; (http://www.fieldtriptoolbox.org/; Oostenveld et al. 2010) toolboxes for MATLAB (2015a, MathWorks). All filtering was performed with a two-pass, Butterworth, fifth order filter. Data were low-pass filtered at 110 Hz, downsampled at 256 Hz, high-pass filtered at 0.1 Hz, re-referenced to the average, divided into 5000 ms epochs (with 1000 ms pre stimulus onset and 1000 ms post-offset) and baseline-corrected relative to the pre-onset interval. Outlier epochs were removed, if the average power over all time samples and channels exceeded 2 standard deviations from the mean over trials; on average, 76% of epochs were retained. Subsequently, data were low-pass filtered at 30 Hz and De-noising Source Separation (DSS; de Cheveigné and Simon 2008; de Cheveigné and Parra 2014) was applied to maximise reproducibility across epochs, keeping the first five components and projecting back into sensor space. Finally, data were averaged over epochs for each channel, condition and subject.

3.4.1.6 Data analysis

For each participant and condition, the root-mean-square (RMS) over channels was calculated at each time sample in the epoch. This was used as a measure of brain activation over time. The distribution of RMS (mean, standard error) was then estimated for each condition using bootstrap resampling across subjects (1000 iterations; Efron and Tibshirani 1993). This was used to calculate the group-level t-statistic of the difference between pairs of conditions at each time-point. T-tests (2-tail) were performed using t-statistics computed on clusters in time, and controlled for a family-wise error rate of 0.05 (Maris and Oostenveld 2007). Additionally, a repeated-measures ANOVA with factors of regularity and alphabet size was performed on the mean RMS power between 1000 and 3000 ms, including all conditions except RAND20 to give a balanced design.

3.4.1.7 Subjects

23 paid subjects took part (mean age 23.3 years, range 20-29 years; 11 female). Two subjects were excluded due to exceptionally noisy EEG data.

3.4.2 Results

Group RMS (RMS of all subjects' RMSs) for the three regular conditions, REG5, REG10 and REG15, alongside RAND20 as a common control, are shown in Figure 3.3a. The brain response shows an N1 peak (at 100 ms post-onset; Figure 3.3d) before rising gradually and reaching a sustained level which persists until stimulus offset. An offset response is visible about 100 ms after sequence cessation. The sustained evoked response is characterised by regular fluctuations at 20 Hz reflecting responses to individual tones. All three REG conditions show an increased sustained response as compared to RAND20. The timing at which the group RMS for REG conditions diverge from RAND20 (taken to reflect the time required by the brain to discover the regularity) increases with cycle length: 406ms [8 tones], 750 ms [15 tones], 1067 ms [21 tones]; for REG with *alph* = 5,10,15 respectively. This is during the second cycle in each case (1.6, 1.5 and 1.4 cycles respectively), before the pattern has repeated completely, although stable statistical significance is reached somewhat later (horizontal lines beneath the RMS plot). As discussed in Barascud et al. (2016), this demonstrates the operation of a rapid, automatic process of regularity detection. Group RMS for REG and RAND of matched *alph* are shown in Figure 3.4. The response to REG is consistently higher than its matched RAND. The scalp voltage map of the difference between REG and matched RAND conditions, calculated between 2.4 and 2.6 s post-onset, is also shown in Figure 3.4. The latency of the divergence between alphabet-matched REG and RAND reflects the time to discover the regular pattern, independently of discovering the frequency pool. For *alph* = 5, means diverge at 421 ms, which is during the 9th tone, reaching significance at 570 ms. For *alph* = 10, the means diverge at 756 ms, during the 16th tone, but only becomes significant much later at 2.66 seconds. REG15 diverges from RAND15 at 968 ms, during the 20th tone, again only becoming significant much later at 1.91 s.

The RMS, over the interval 2000-3000 ms post-onset, extracted from each condition, was subjected to a repeated-measures ANOVA with regularity (REG vs. RAND) and *alph* (alphabet size of 5, 10 or 15 tones) as within subject factors. This yielded significant main effects of regularity ($F_{1,20} = 29.9$, $p < 0.0001$) and of *alph* ($F_{2,40} = 3.68$, $p = 0.034$), with no interactions. Post-hoc comparisons on the main effect of *alph* revealed significant differences only between *alph*=5 and *alph* = 15 ($p = 0.017$). The relative mean increase between RAND and REG of matched *alph* was 29.7%, 36.7% and 27.5 % for *alph* = 5, 10 &15 respectively.

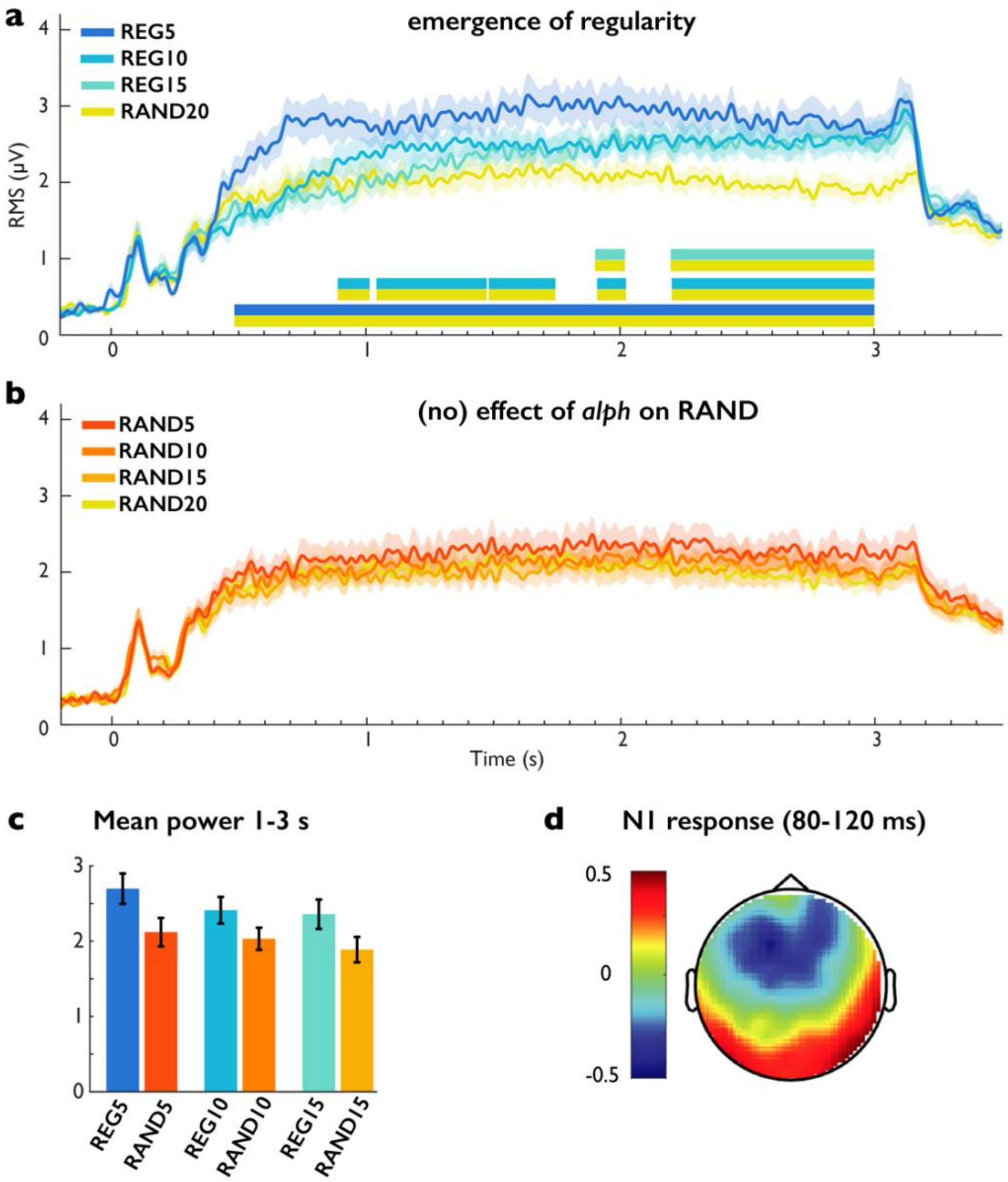

Figure 3.3 Sequence-evoked responses

EEG sequence-evoked responses for Experiment 1. Shaded error margins show ±2 SEM. **a:** EEG evoked responses (group RMS over all channels) for RAND20 and REG of different alph, over the entire epoch. Horizontal bars below plots indicate time intervals where cluster-level statistics showed a significant difference between each of the REG conditions and RAND20, colour of the bars denotes which two conditions are compared. **b:** Comparison of all four RAND conditions with different alph. There were no significant differences between RAND20 and RAND5, 10 or 15. **c:** RMS average power over 1000-3000 ms. There is a main effect of alph and of regularity with no interaction. **d:** N1 onset response (80-120 ms) to all conditions.

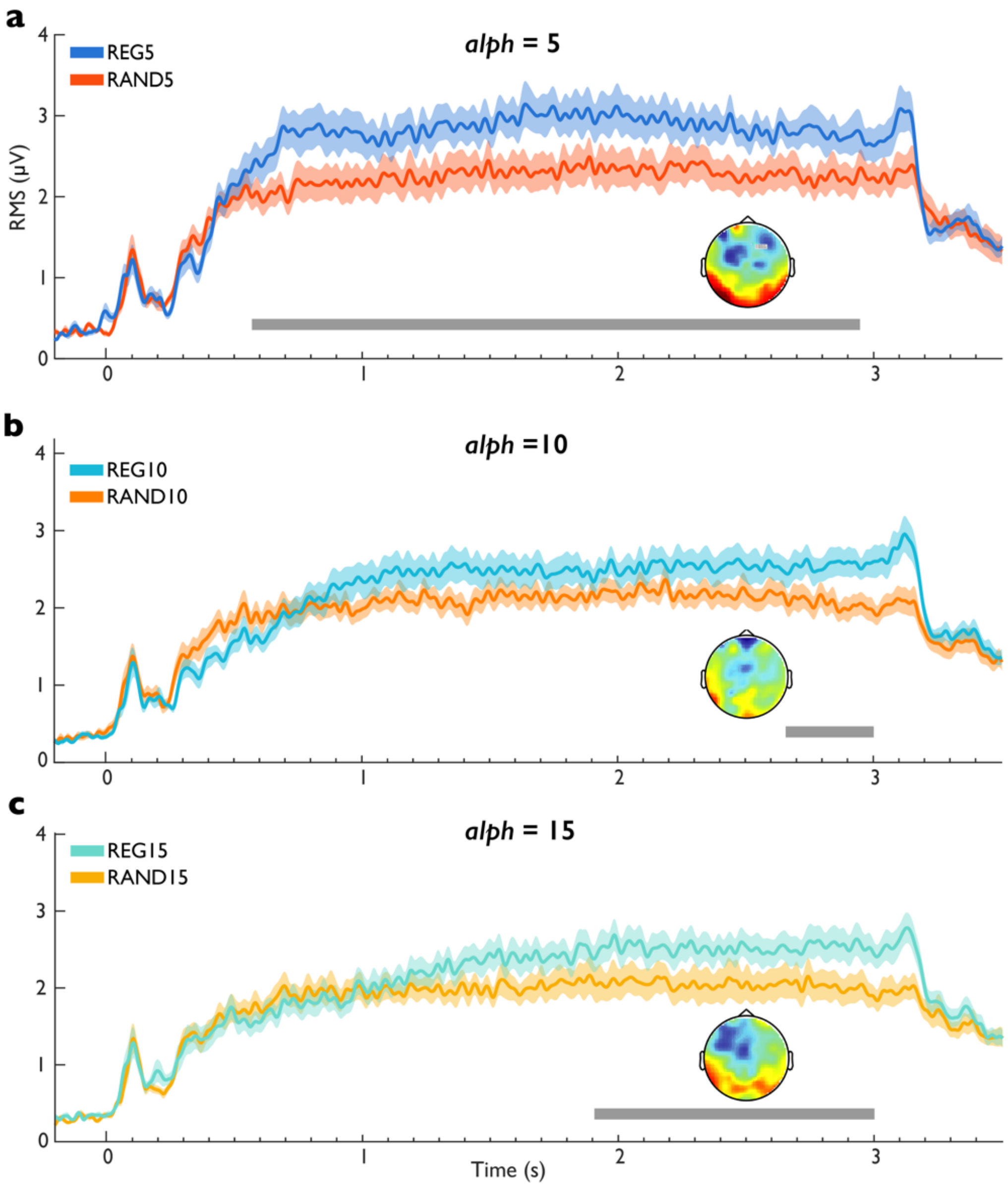

Figure 3.4 Sequence-evoked responses
EEG sequence-evoked responses for Experiment 1. The responses to REG are re-plotted alongside their respective RAND controls; for **a:** alph = 5, **b:** alph =10, and **c:** alph =15. Horizontal grey bars below plots indicate time intervals where cluster-level statistics showed a significant difference between each pair of conditions. The scalp voltage map of the difference between each pair of REG and RAND conditions, calculated between 2.4 and 2.6 s post-onset, are plotted inset below each RMS plot.

3.4.3 Discussion

Experiment 1 demonstrated a rapid differentiation in brain response between regular and random sequences, replicating the MEG results (Barascud et al. 2016). However, the EEG responses are somewhat noisier than those measured with MEG, likely influenced by a number of factors including ambient electric noise and the smaller number of sensors used here (64, versus 274 in MEG) which impairs the efficiency of de-noising. Electrode voltage drifts, which introduce low frequency noise, may also have affected the robustness of the observed sustained responses.

Overall, the data demonstrate that regular, predictable sequences lead to a dramatic power increase of some 30% which is remarkable for evoked responses, and suggests a large change in the underlying neural activity. This finding is surprising in light of previous work consistently reporting reduced evoked responses to predictable patterns and interpreted as reflecting reduced prediction error (Costa-Faidella et al. 2011); for a review see (Baldeweg 2008). The discrepancy with previous work may be due to much of the existing work using sound patterns which confound repetition with predictability, making it difficult to dissociate adaptation effects from those purely due to prediction. The present paradigm, using wide-band signals and complex sound patterns, allows us to control for simple effects of adaptation (see additional discussion in Barascud et al. 2016). Furthermore, we use very rapid sequences, where the perception of patterns pops-out spontaneously rather than being consciously trackable. It is likely that the neural processes involved in extracting the regularity are different from those implicated in work using slower temporal patterns (e.g. Hsu et al. 2014; Auksztulewicz and Friston 2015), which allow high-level (conscious or mnemonic) prediction of future events.

One possible explanation for the activation pattern observed here is that it reflects automatic, bottom-up driven attentional capture by REG patterns. This attentional process will be the focus of the rest of this paper. The behavioural experiments below investigate the hypothesis that the large, sustained amplitude shift which was observed for the REG stimuli may reflect increased perceptual salience (Itti and Koch 2000; 2001). That is to say, the more reliable REG stimuli trigger an automatic (exogenous) attentional bias. This proposition leads to the testable prediction that REG and RAND will have different effects on behaviour, reflecting an attentional bias towards regularity in REG sequences, even when task-irrelevant.

3.5 Experiment 2

This experiment aimed to measure the (assumed) behavioural consequences of attentional capture by regular sounds. We evaluated performance on a demanding listening task, with REG or RAND sequences presented concurrently, as task-irrelevant distractors. If REG patterns spontaneously capture exogenous attention, we predicted that REG sequences will prove more detrimental to performance than RAND sequences. This prediction is in line with previous behavioural experiments, whereby task-irrelevant stimuli outside the focus of attention can result in attentional capture manifest as degradation in performance in a behavioural task (Escera et al. 1998; Cervantes Constantino et al. 2012; Sohoglu and Chait 2016b).

The main task was based on an auditory change-detection paradigm (Cervantes Constantino et al. 2012; Sohoglu and Chait 2016b). Stimuli were artificial acoustic scenes, comprised of multiple simultaneous tone-pip streams, each characterized by a distinct, constant frequency and amplitude-modulation (AM) rate. The task required listeners to detect occasional changes (appearance or disappearance of one stream) within these scenes. This simulates the challenges faced by listeners in natural acoustic scenes, in which many concurrent sound sources must be processed and monitored simultaneously.

We presented the change detection task and REG/RAND distractor sequences concurrently to different ears, such that they competed directly throughout the trial (Figure 3.5a). If REG patterns spontaneously capture exogenous attention, we predicted that REG sequences will prove more detrimental to performance than RAND sequences.

3.5.1 Methods

3.5.1.1 Stimuli

The stimuli and experimental approach for the change detection paradigm (Figure 3.5a) are described in detail in a previous study (Cervantes Constantino et al. 2012). In brief, stimuli were artificial acoustic scenes consisting of 8 concurrent streams of tone pips, each with a unique frequency (between 200 and 4000 Hz) and AM rate (3 to 35 Hz). 50% of the stimuli contained a change partway through the scene: appearance (CA) or disappearance (CD) of a stream. Scene changes occurred between 1000 and 2000 ms post-onset. The overall stimulus duration was between 2000 and 4000 ms.

On each trial, a scene stimulus and a REG10 or RAND10 distractor sequence (with equal probability) were presented concurrently at the same dB level, to different ears, such that they competed directly throughout the trial. Each REG/RAND sequence consisted of between 40 and 80 tones. For REG sequences this constituted between 4 and 8 cycles; i.e. sufficient for the regularity to become perceptually established.

3.5.1.2 Procedure

Stimuli were blocked by change type (CA or CD), with 50% of the trials in each block (160 overall) containing a change. The ISI was randomized between 700 and 2000 ms. REG or RAND sequences were randomly paired with each scene stimulus. Subjects were instructed to attend to the ear containing the scene and respond by button-press as soon as they heard a change. To avoid confusion, the ear of presentation was fixed throughout the experiment, but counterbalanced across subjects. Subjects were naïve to the structure of the REG/RAND sequences, and told these were simply distractors to the main change-detection task. The block order was counterbalanced between subjects, and a break was allowed after every 40 trials. The session began with a short training block where feedback was given on each trial.

3.5.1.3 Subjects

10 subjects participated in this experiment (mean age 24.0 years; 7 female).

3.5.2 Results

Figure 3.5b,c shows reaction times (RT) and sensitivity (d'; Tanner and Swets 1954) scores for detection of scene changes with REG or RAND distractors presented concurrently. A repeated-measures ANOVA was performed on RT and d'; with change type (CA/CD) and distractor regularity (REG /RAND) as factors. RT showed main effects of change type ($F = 16.299$; $p = 0.003$) and regularity ($F = 8.064$; $p = 0.019$) with no interaction. Similarly, d' showed a main effect of change type ($F = 18.244$; $p = 0.002$) and regularity ($F = 9.786$; $p = 0.012$) with no interaction.

The results reveal that, contrary to our hypothesis, RAND is more detrimental to performance than REG. The data are consistent with the interpretation that RAND is harder to ignore, at least when in direct competition with a concurrent, task relevant, auditory stream (see also Andreou et al. 2011, and discussion in §3.7.2).

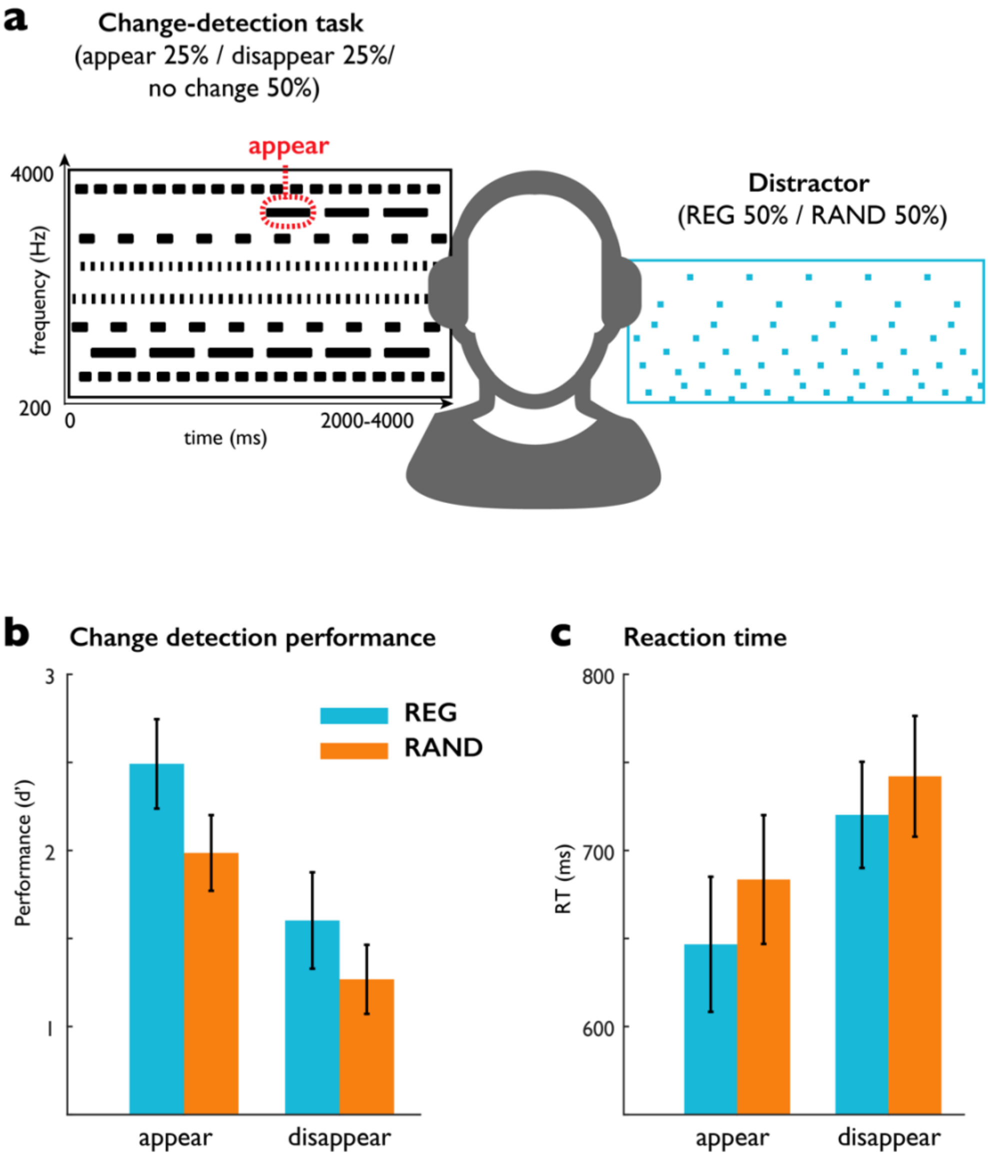

Figure 3.5 Experiment 2

a: A schematic representation of the stimulus paradigm. Scene stimuli and REG or RAND sequences (RAND in this example) were presented concurrently to different ears (counterbalanced across subjects). All scenes contained eight streams. This example shows a scene with an appearing stream, which is indicated in red. **b:** Behavioural results; performance (indexed by d') and **c:** reaction time for detection of change events. Error bars show ±1 SEM.

3.6 Experiment 3

Rather than using REG or RAND sequences as task-irrelevant distractors, Experiment 3 placed REG and RAND sequences in direct competition as task-relevant streams. Here REG and RAND were presented concurrently and both actively monitored for targets (a silent gap). This is in contrast to Experiment 2, where performance on the task required ignoring REG or RAND stimuli. We predicted that, when both sequence types are monitored simultaneously, gaps in REG sequences should be more readily detectable than gaps in RAND sequences. This design is similar to that used in Zhao et al. (2013), who demonstrated that targets embedded within regularly repeating visual streams are more easily detected, even though the regularity of the stream was not itself goal-relevant.

3.6.1 Methods

3.6.1.1 Stimuli

This experiment used REG5 and RAND5 sequences consisting of 50-ms tone pips interspersed with 50-ms gaps. Trials involved the presentation of two concurrent sequences, one in each ear (Figure 3.6). Sequences could be both REG or both RAND, or one of each. In order to facilitate the perception of the two sequences as independent concurrent streams, the sequences were staggered by 50 ms such that tones occurred in alternation between the ears. In addition, sequences were spectrally separated such that the sequence in the right ear was always a higher pitch. The tones were chosen from a pool of 13 logarithmically-spaced frequencies between 1587 and 6205 Hz for the right ear and between 280 and 1122 Hz for the left.

On 50% of trials, one of the sequences contained a target (an omission of two consecutive tones). Stimuli were 6000 ms long. When present, the target occurred at least 2000 ms after stimulus onset (following 4 REG cycles); i.e., at a point in the REG stimulus when the regular pattern has been established.

The main experiment consisted of the following conditions: (ii) RAND sequences in both ears (**RAND-RAND**, 25% of trials); (ii) REG sequences in both ears (**REG-REG**, 25% of trials); (iii) REG sequence in one ear and RAND in the other (**REG-RAND**, 50% of trials). The target occurred in one of the two sequences with equal probability. For REG-RAND, we denote the stream containing the target using the prefix **t** for 'target', and the other sequence with the prefix **d** for 'distractor'; thereby sub-dividing this condition into (iiia) **tREG-dRAND** and (iiib) **dREG-tRAND**. Each condition was counterbalanced across the two ears such that target occurrence and REG versus RAND sequences were equally likely in each ear. Before the main session subjects also completed a block where only a single sequence (REG or RAND) was presented to one of the ears with equal probability; a target was present on 50% of trials. These conditions are denoted as REG- and RAND-.

3.6.1.2 Procedure

Subjects were instructed to press a keyboard button as soon as they heard a gap in a sequence. Feedback was provided after each trial. The single-sequence block contained 80 stimuli (40 for each of REG- and RAND-). The main experiment contained 120 trials for each of the four dual-sequence conditions, presented in a randomised order. Subjects were given short breaks every 10 minutes.

3.6.1.3 Subjects

11 subjects participated in this experiment (mean age = 25.7; 7 female)

3.6.2 Results

We used hit rates as our primary outcome measure, as most subjects produced low false positive rates (2.1% for single-sequence, 2.0% for dual-sequence) yielding 'artificially' high d' scores. Hit rates were deemed the most unambiguous and representative outcome measure of performance in this task.

Figure 3.6b,c show the hit rate for all conditions, separated by whether the target is in the REG or the RAND stream, for both single-sequence and dual-sequence stimuli. The hit rates for targets in the single sequence (REG- and RAND-) conditions are the left-most bars in each plot. In the single-sequence condition, subjects were better at identifying targets in regular streams compared to random streams (F = 20.6; p = 0.001). These findings are consistent with previous work showing that performance is improved when targets are embedded in temporally regular sequences (Barnes and Jones 2000; Jones et al. 2002; 2006). This is the case even when the dimension in which the regularity expressed is independent of the dimension along which targets differ.

In order to test the main hypothesis that regular sequences would 'pop-out' and attract attention, we compared performance for dREG-tRAND and tREG-dRAND, as both contain simultaneously presented random and regular sequences (Figure 3.6d). We postulated that regularity would bias attention, leading to improved performance when targets were embedded in regular streams (tREG-dRAND) as opposed to random (dREG-tRAND). A repeated-measures ANOVA showed no significant difference between the average hit rate values for tREG-dRAND and dREG-tRAND (hit rates were 0.67 and 0.65 respectively; F = 1.2; p = 0.3). These results suggest that regular sound patterns do not bias attention.

A repeated-measures ANOVA was conducted on hit rates in the dual-sequence conditions, with factors for regularity of the target stream (REG/RAND) and the parallel stream (REG/RAND). There was a main effect of target stream (F = 51.48; $p < 0.01$); here again subjects were overall better at detecting targets in REG. We also found a main effect of parallel stream (F = 26.97; $p < 0.01$); revealing overall poorer target-detection when the parallel stream was RAND relative to when it was REG. This pattern is in agreement with the outcomes of Experiment 2, and consistent with the interpretation that RAND patterns incur increased demand on processing resources (discussed further below). However, there was no interaction between the two factors, suggesting that a RAND parallel stream incurred an equal cost to target detection performance in a RAND or a REG stream.

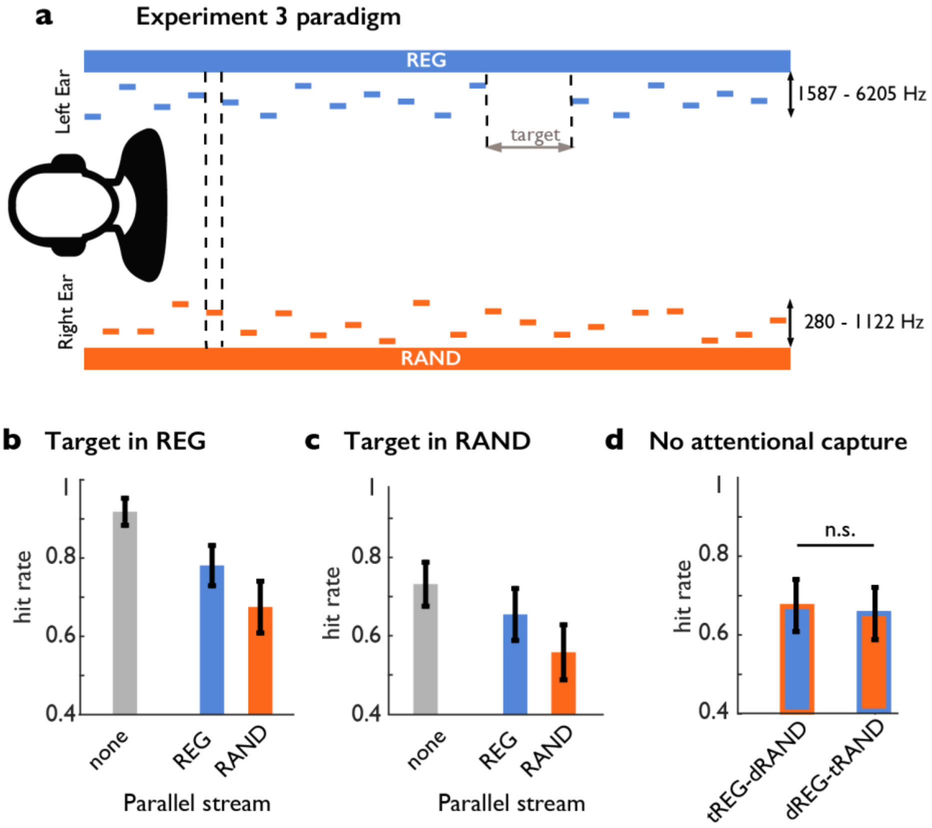

Figure 3.6 Experiment 3

a: A schematic representation of the stimulus paradigm. On each trial, two concurrent sequences (each either REG or RAND) were presented, one to each ear. Individual tones were interleaved as demonstrated with the black dashed line. On 50% of the trials, one of the sequences contained a 200 ms gap (target, shown in grey, top). The sequences presented to the right ear were of a higher pitch than to the left, such that the frequency ranges did not overlap. **b:** Hit rate for all conditions with a target in the REG stream; single-sequence REG (left) and dual-sequence (right). **c:** As for b; except all conditions had a target in the RAND stream. Error bars show ±1 SEM. **d:** Hit rate re-plotted for target detection when REG was presented to one ear and RAND to the other. The sequence containing the target is indicated with the prefix t for ‘target’, the other stream is denoted with the prefix d for ‘distractor’.

3.7 Discussion

Brain responses measured with functional magnetic resonance imaging (fMRI), MEG (Barascud et al. 2016) and EEG (as seen here) show consistently increased activation to regular acoustic patterns, relative to matched random stimuli. One interpretation of these systematic, pronounced effects is that they indicate large differences in attentional capture between regular and random patterns, such that regular patterns automatically and involuntarily attract more attention. This account of the imaging data is consistent with previous behavioural work in the visual modality (Zhao et al. 2013) and is broadly in line with the fact that sensitivity to predictable patterns in the natural environment is a major pre-requisite for survival. Organisms produce regular, periodic motor sequences, such as locomotion and vocalizations, which are expressed as a pattern in the temporal succession of sensations. The ability to automatically orient towards such patterns within a crowded, noisy scene is often critical for continued existence.

3.7.1 Attentional capture by regularity?

The behavioural experiments reported here aimed to identify a behavioural correlate for the observed brain response effects. Using two tasks designed to probe different aspects of attentional capture, we consistently find no evidence for the exogenous capture of attention by regular acoustic patterns. Despite the sizeable change in the EEG signal associated with processing regular (REG), relative to random (RAND) tone pip patterns, REG sequences were not more distracting than matched random sequences when task irrelevant, and were also no more perceptually salient when participants were actively monitoring REG and RAND streams concurrently.

Whilst there is no evidence for REG sequences being more perceptually salient than RAND sequences (or vice versa), Experiments 2&3 suggest that RAND sequences are more computationally demanding, and hence more distracting, than REG sequences (e.g. Falk and Konold 1997; the discussion below disentangles these issues).

The paradigm in Experiment 2 shared key similarities with the EEG experiment (Experiment 1). Participants were naïve to the nature of the distracting REG or RAND patterns and focused on a different task. A change detection task, rather than a visual task similar to that in the EEG experiment, was chosen because: (i) its rapid nature allowed us to probe behaviour more frequently, and hence efficiently; and (ii) a competing auditory (rather than a visual) task is more likely to reveal effects of distraction, because it poses more competition for shared resources (see review in Chait et al. 2012). It is therefore unlikely that failure to observe effects of attentional capture is due to the difference in task per se. Furthermore, by removing the decoy task altogether, Experiment 3 constitutes a stricter test for a possible attentional bias. When REG and RAND are monitored concurrently, we observe equal gap-detection performance whether the target is in REG or RAND, which suggests that they do not differ in their perceptual salience.

The reasons for the discrepancy with results from vision (Zhao et al. 2013), where effects of attentional capture by regularity have been reported, is unclear and may be due to many factors, perhaps including a genuine difference in the mechanisms of attentional allocation in the visual and auditory domain. Further work, directly comparing the two modalities, is required to resolve this issue.

3.7.2 Processing of regular vs. random sequences

The results of Experiment 3, demonstrating increased sensitivity to targets in REG relative to RAND sequences when presented alone, are consistent with many previous demonstrations that regularity facilitates behavioural performance. These studies, albeit mostly using regularity in the temporal dimension rather than in frequency as we do here, consistently show that regularity facilitates behavioural performance. Expected events are detected and assessed more rapidly and accurately than unexpected events (Barnes and Jones 2000; Jones et al. 2002; Correa and Nobre 2008; Bendixen et al. 2009; Leaver et al. 2009; Ellis and Jones 2010; Jaramillo and Zador 2011; Geiser et al. 2012; Rohenkohl et al. 2012; Lange 2013). This occurs, as is the case here, even when the task dimension is orthogonal to the feature dimension over which the regularity is defined (e.g. Jones et al. 2006) and hypothesised to arise due to the ‘pre-activation’ of the relevant neural machinery for processing predicted events (Hughes et al. 2001; Bendixen et al. 2009).

The same processes have been demonstrated to contribute to the suppression of regular streams when they are not behaviourally relevant. For example Andreou et al. (2011) demonstrated that it is easier to ignore a temporally regular sequence, relative to a temporally irregular sequence (Bendixen et al. 2010; Devergie et al. 2010; see also Rimmele et al. 2012). Similarly, in Experiment 2, we show that REG sequences are less distracting than RAND sequences when participants are required to ignore those sequences and focus on a competing change detection task. A potential mechanism for this effect is supplied by predictive coding (Clark 2013; Hohwy 2013), whereby predictable inputs are attenuated by top-down predictions, and the resulting prediction error triggers a process of updating the internal predictive model. Regularity allows the derivation of a predictive rule, therefore it becomes easier to ‘explain away’ the irrelevant stimulus, by suppressing the prediction error with a closely-matching top-down prediction. Irregular stimuli demand more resources for processing as they elicit a constant stream of prediction errors and thus constantly trigger model updating. This may be taken to suggest that (unpredictable) RAND sequences are more perceptually salient. From the point of view of predictive coding, this is sensible because RAND sequences are characterised by higher information content than REG sequences. Friston et al. (2012) define salience in terms of the ability to reduce uncertainty or to inform hypotheses about the sensory scene being sampled. In the visual domain, this is usually measured in terms of Bayesian surprise (Itti and Baldi 2009) and more generally as information gain or epistemic value (Friston et al. 2015). However, whether the theoretically information-rich RAND signals are useful in reducing uncertainty about high level representations is an open question. In other words, is unpredictability itself salient? The lack of a bias in performance between REG and RAND sequences presented in direct competition (Experiment 3) suggests not.

3.7.3 If not attentional capture, what is the source of the EEG effect?

EEG and MEG measurements demonstrate an increase in the amplitude of the sustained response for REG relative to RAND stimuli. In fMRI this is associated with greater activation for REG relative to RAND sequences across a large portion of the superior temporal gyrus, including Heschl's gyrus and planum temporale (Barascud et al. 2016). The behavioural results reported above suggest that this increased activation is not associated with attentional bias towards (or increased perceptual salience of) REG sequences.

Critically, the above explanation for why RAND sequences were, in some cases, more detrimental to performance than REG sequences implies *more* activation (increased demand for computational resources) for random patterns relative to regular ones. This may seem contradictory to the brain-level effects. However, it is possible, that the increased auditory cortical activation for regular patterns observed in M/EEG and fMRI reflects increased *inhibition.* It is difficult to dissociate excitatory and inhibitory activation with non-invasive brain imaging techniques, rather future computational and electrophysiological tools would be critical for exploring this possibility. Indeed, recent findings in animal models demonstrate a critical role for inhibition in shaping the response of primary auditory cortex neurons to regularly repeating sounds in the context of an oddball paradigm (Natan et al. 2015).

Another potential explanation for the larger response to REG is that regularity detection is associated with heightened sensitivity (increased gain) of the sensory units activated by the regular pattern. According to this *precision-weighting* account, precise, i.e. highly predictable, sensory streams are preferentially weighted by increasing the post-synaptic gain of the relevant (prediction error) units (Friston 2005). Importantly, this can occur within the remit of automatic processing, so doesn't entail attentional capture (Feldman and Friston 2010).

Lastly, it is possible that the increased sustained response we observe is due to another process (or indeed multiple processes) such as learning, working memory or recognition of a match to a memory of previous stimulation (see Barascud et al.

2016 for further discussion). This interpretation is consistent with the diffuse source network including auditory cortex, hippocampus and inferior frontal gyrus identified by Barascud et al. (2016) as contributing to the brain response to structured sequences.

To summarise, a picture emerges from these results in which regularity in non-attended items does not capture attention. In fact, as demonstrated in Experiment 2, it is random stimuli which can be more distracting than regular ones. Consistent with the literature, we found that regularity does, however, aid in scene analysis by being easier to ignore (Experiment 2) and requiring fewer resources to process (Experiment 3). Collectively the behavioural and brain imaging findings can be reconciled by considering both to result from mechanisms that minimise surprise and uncertainty about the world (Schröger et al. 2013; Bendixen 2014).

3.8 Addendum

3.8.1 EEG methods & results

This chapter presents work carried out and published at the start of the PhD. In later experiments, techniques for analysing the EEG data to better account for slow drift in the signal were developed (see §2.3) and applied in the pre-processing stages in later chapters. Because the present chapter has already been published, the decision was taken not to re-analyse the EEG results in Experiment 1.

However, it has since become clear that the pre-processed data are noisier than in several subsequent chapters. For instance, there was no difference between RAND5 and RAND20 at any time during the sustained response. However, the RMS amplitude was significantly higher for RAND5 than for RAND20, and for RAND15 relative to RAND20, in Barascud et al. (2016). Further, in Chapter 4, RAND5 evokes significantly higher response power than RAND20, and this effect is present for much of the sequence (see Figure 4.3). REG10 and RAND10 show a significant difference in the present chapter only briefly, from 2.66 seconds after onset, whilst the means diverge much earlier, at 750ms, which is a similar latency to the mean divergence time seen in Chapter 5 as well as the MEG study (Barascud et al. 2016).

3.8.2 Discussion of relevant recent papers

Recently, more evidence has come to light that predictability does not capture attention. Meijs et al. (2018) used a visual cue-target paradigm to determine whether predictable stimuli reduce (as shown here) or increase (as shown by Zhao et al. 2013) attentional capture. In Meijs et al.'s study, cues could appear at one of two spatial locations, and could validly or invalidly predict the location of a subsequent target, with equal probability. When the cue and target mismatch, reaction times are expected to be slower than when they match: this difference is termed the validity effect, and indexes how distracting the (irrelevant) cue location is to detecting the target. Manipulating the prior predictability of the cue location, by biasing the cue to occur on one side on 80% of trials, did not modulate the validity effect, leading the authors to conclude that prior information on the features of the distractor does not modulate attentional capture. However, Experiment 2 in this chapter found that distractor predictability was indeed advantageous to target detection, which was interpreted as regularity being easier to ignore or suppress. The discrepancy could be down to any of several differences between the paradigms. Exogenous attentional capture can be measured in many ways, and different tasks present different demands on different perceptual systems. Suppressing the distractor was arguably advantageous in the behavioural tasks presented in this chapter, because the target and distractor were never presented in the same ear, and it is known that humans are capable of strongly inhibiting competing information presented dichotically (Broadbent 1952; i.e. different stimuli delivered to each ear: Cherry 1953). However, as the authors themselves discuss, it would be disadvantageous to suppress the spatial location of the cue, even though its location does not provide information on the target location, because the target may occur in that same spatial location and only after a short delay. The present Experiment 3 measured exogenous attentional capture using a different paradigm, and, like Meijs et al. (2018), also found no evidence for attentional bias towards or away from predictable features, when regular and irregular stimuli were presented simultaneously. In contrast to Zhao et al. (2013), stimulus predictability was shown not to attract visual attention when irrelevant (Alamia and Zenon 2016). The experimental design distinguished predictive stimuli (from which future stimuli could be predicted) from predicted stimuli, and neither attracted (visual gaze) attention except for when the predictable relationship pertained directly to the task target.

Predictable stimuli have even been found to be less distracting, or less salient, in several recent studies; in line with the results of Experiment 2. A behavioural and eye-tracking study provided further demonstration that repetitive stimuli do not capture attention (Huang and Elhilali 2017). Naturalistic auditory scenes were presented dichotically whilst participants continuously indicated which stream captured their attention more. Identical scene components resulted in a lower salience rating on subsequent repetitions. In an audio-visual study, Xie et al. (2018) found unpredictable sequences of auditory stimuli to slow reactions to visual targets when compared to the same auditory stimuli presented in repetitive sub-sequences.

Chapter 4. Predictability and Repetition

4.1 Summary

This chapter expands on the work presented in Chapter 3, using EEG to measure the response to complex tone-pip sequences (REG and RAND) presented to distracted subjects. Simpler regularities are used than in the previous chapter, in order to better link the sustained response phenomenon to existing literature.

In classical studies of auditory pattern learning, responses exhibit repetition suppression, where sequences of repeated tones show a reduced neural response. This may be due in part to adaptation, but also may indicate suppression of expected stimuli. This is thought to form a building block of regularity learning.

However, more complex patterns used in this thesis show the opposite effect, where predictable (REG) patterns distributed over a range of frequencies show a strongly enhanced brain response, compared to frequency-matched random (RAND) sequences. This effect cannot be explained by neuronal adaptation, and sits at odds with reduced evoked responses to predicted stimuli more generally.

This experiment reconciles auditory sequence predictability and repetition in a single paradigm, by incorporating a condition where sequences consist of exactly repeating tones at a single frequency, alongside more complex REG and RAND. Our results indicate a system for automatic monitoring of predictability in the auditory environment which is distinct from, but concurrent with, repetition suppression. Further, it is suggested that exact repetition of a single frequency does not constitute a good model for regularity from which generalisable conclusions can be drawn.

4.2 Statement of Contribution

The work in this chapter was undertaken with a Masters student (Candida Tufo) under my supervision. Maria Chait and I designed the experiment. CT collected the data. I wrote the code for stimulus presentation, and analysed the data. MC and I interpreted the data. I wrote this chapter.

4.3 Introduction

4.3.1 Motivation

The EEG results in Chapter 3 replicate previous work (Barascud et al. 2016) in showing that predictability and response power over the whole scalp are related for rapid, passively-perceived sound sequences. However a divide exists between these results and previous literature on the direction of this relationship. Decreased responses to repeated stimuli (Haenschel et al. 2005; repetition suppression; Grill-Spector et al. 2006) are seen in a wide range of sensory modalities and imaging methods (Desimone 1996; Summerfield et al. 2008; de C Hamilton and Grafton 2009; Auksztulewicz and Friston 2016). Reduced responses are also seen the more generally predictable a stimulus is, conceptualised by predictive processing accounts as a manifestation of reduced prediction error (Friston 2005). In fact, repetition suppression is often cited as evidence for predictive coding accounts (Baldeweg 2006; Garrido, Kilner, Kiebel, et al. 2009; Auksztulewicz and Friston 2016) which consider repetition suppression as the result of regularity detection and subsequent ‘explaining-away’ of the repeated input.

4.3.1.1 Repetition as a model of predictability

As a model of predictability, stimulus repetition is potentially problematic in that the observed effects may not be due to the formation of predictions, but could instead reflect fatigue of neuronal populations (Budd et al. 1998; Grill-Spector et al. 2006); without invoking anticipatory, top-down suppression informed by an internal model of the regularity (but see e.g. Summerfield et al. 2008). Perhaps more importantly, even if repetition suppression is not merely a result of neuronal adaptation, it represents only a very basic predictability rule, and it is not a given that findings using repetition paradigms would generalise to more complex forms of regularity. Alternative procedures have been designed to dissociate the effects of repetition and expectation, (Todorovic et al. 2011; Todorovic and de Lange 2012), finding that indeed reduced responses occur to predicted tones, regardless of whether the predictability is manipulated through repetition.

4.3.1.2 REG and RAND as a model of predictability

In Chapter 3, two forms of predictability were manipulated, both serving to increase evoked responses. A potential factor in this discrepancy, relative to previous literature, is the nature of the stimuli and analyses used in the present work. The sequences are much faster and the tone pips much shorter than are typically used for auditory evoked potential studies, which often employ tone durations, and silent intervals between tones, of hundreds of milliseconds. This could contribute to altering the observed effects in several ways. As the tones I use are presented without any silent ITI and with such a short SOA of 50 ms, the cortical evoked responses to each tone overlap with the next. Disentangling the influence of these overlapping but time-shifted responses is a difficult problem, particularly if the responses to individual events aren't linearly summated (e.g. as discussed in Crosse et al. 2016). Additionally, the evoked response measure used here is a slow, DC-like shift which accumulates over several tones. This is only visible if timeseries data are pre-processed without any high-pass filter or with a very low cut-off; whereas a filter cut-off of 1 or 2 Hz is commonly applied at the start of the pre-processing pipeline. Thirdly, auditory studies exhibiting repetition suppression mostly use exact repetition of tone pips at a single frequency as a model of stimulus regularity. REG sequences are more complex, covering a wide bandwidth of frequencies, so the repeating section itself is more intricate. The regularity effect is present when comparing REG and RAND with the same frequency content, thus controlling for neuronal adaptation. Using these stimuli may therefore tap different processes than the classical auditory studies, despite both types of stimuli being employed to model the effects of predictability. For all of the above reasons, it is difficult to relate the sustained response to REG and RAND to findings from commonly-used paradigms in the field.

To reconcile these opposing effects of predictability on auditory responses, in addition to 'standard' REG and RAND sequences, we included in this experiment increasingly simple regularities including sequences consisting of a single repeating tone. RAND patterns in the present experiment had an alphabet *alph* of 3, 5 or 20 unique tones, shuffled in equal quantities. REG used an *alph* of 1, 3 or 5. Shorter regularity cycles were included, allowing comparison to the results in Chapter 3 whilst expanding the range of stimuli to incorporate ever-simpler repetition. Taken to the extreme, the shortest cycle (REG1) implements simple repetition as used in the literature demonstrating stimulus-specific adaptation, repetition suppression and adaptation of the standard in MMN paradigms. However the temporal structure of REG1 was matched to the other REG and RAND stimuli, permitting more direct comparison of the regularity enhancement seen in the sustained response, and repetition suppression.

From the repetition suppression literature, the evoked response to REG1 should show evidence of decreased responses when compared to less repetitive stimuli. However, REG1 is highly predictable; according to an alternative account arising from Chapter 3, this should serve to increase the sustained response above a random, less predictable sequence.

The EEG responses to these stimuli were analysed in a time-resolved manner, focussing separately on the onset, sustained and offset portions of the evoked response. This approach increases sensitivity to potentially opposite effects occurring at different times. The onset responses likely reflect initial processing in auditory cortex. The sustained response is expected to reflect more complex processing. The overall sustained response power has previously been shown to follow stimulus predictability. Due to the cyclical nature of the regularity used here, it is hypothesised that this regularity itself will be visible as a peak at the cycle rate in the brain response, on each trial. Such an effect was indeed seen in local field potential (LFP) recordings in non-human primates in response to repetition of spectral patterns in tone pip streams (Barczak et al. 2018), using similar stimuli to the present thesis. The strength of the response to individual tones was separately estimated using spectral power at the tone rate; to provide an alternative measure of gain on auditory responses. This measure is again expected to be correlated with sequence predictability. Sequence offset is anticipated to reveal the statistical properties of the preceding sequence (Andreou et al. 2015). Specific hypotheses for particular conditions are set out in more detail below.

4.3.2 Hypotheses

4.3.2.1 Regularity increases evoked responses

Firstly, replication of the effects seen for REG5 and RAND5 in Chapter 3 and in MEG (Barascud et al. 2016) is expected: namely, that REG will rise above RAND in the sustained response power, diverging after a cycle plus four tones; i.e. at around 450 ms, then remaining at a higher level for the remainder of the sequence. It is predicted that REG3 will show a similar increase in the evoked response relative to RAND3, tending to the same level as REG5, but with an earlier divergence time. Based on an ideal observer model it is suggested that this divergence would occur at a cycle length plus 4 tones, which corresponds to 350 ms (see Figure 4.1b). A less precise expression of this prediction is that REG3 will diverge from RAND3 before REG5 diverges from RAND5. Finally, offset responses are expected to be greater the more predictable the sequence; i.e. greater for REG than for RAND.

4.3.2.2 Regularity increases induced responses at the cycle rate

As REG contains a repeated cycle, at a periodicity of 150 or 250 ms for *alph* = 3 and 5 respectively, it is expected that this will be detectible in the brain response as a peak in the frequency spectrum at 6.67 or 4Hz (respectively). However, as the repeated chunk is different on each trial, the phase of this response can’t be expected to align across trials, as such the power spectral density at the frequencies-of-interest will be computed on a single-trial basis. This would be expected to be absent for matched RAND; forming the condition against which the presence of a cycle-rate oscillation will be assessed. As REG1 does not have a matched RAND condition, this effect will only be assessed for *alph* = 3 and 5.

4.3.2.3 RAND response power is inversely related to *alph*

Second, the sustained responses to the three RAND conditions will be modulated by alphabet size such that the sustained response magnitude will be RAND3>RAND5>RAND20; replicating the effect seen in MEG (Barascud et al. 2016).

4.3.2.4 REG1 behaves like REG3 and REG5

REG1 is the key condition where predictions informed by the literature and by the results in Chapter 3 differ. Based solely on the explanations put forth in the discussion of Chapter 3, it is hypothesised that the sustained response to REG1 will be of similar magnitude as compared to REG3 and REG5, because all are deterministic, and have a similar information content per tone (Figure 4.1b). However it will rise to this level sooner, as the regularity begins from the second tone, as opposed to the fourth or the sixth for REG3 and REG5 respectively. The response will then remain flat until offset. An alternative hypothesis is that the REG1-evoked response will exhibit adaptation, manifest as a decrease of the sustained response throughout the trial.

4.4 Methods

Electroencephalography (EEG) was used to record the evoked responses to regular (REG) and random (RAND) sequences of tone pips, varying the size of the frequency 'alphabet' used. The participants were naïve to the experimental manipulations, and were asked to ignore the sounds; watching a silent, subtitled film of their choice throughout.

4.4.1 Stimuli

As for Chapter 3, stimuli were 3000-ms sequences consisting of 60 tone pips (50 ms each) with frequencies drawn from a pool of 20 logarithmically-spaced values (222 to 2000 Hz; increasing by 12% at each step). Sequences were either regular (REG) or random (RAND) and were generated anew for each trial with Matlab (R2015a; Mathworks).

Figure 4.1a shows example spectrograms for each stimulus condition. For the REG conditions, a number *alph* of frequencies (1, 3 or 5 for **REG1**, **REG3** and **REG5** conditions respectively) were randomly chosen from the frequency pool, arranged in random order and then repeated for the duration of the sequence, yielding a regularly-repeating pattern. Matched **RAND3** and **RAND5** sequences were generated by shuffling the REG conditions with *alph* = 3 and 5. **RAND20** was a randomised sequence using the full frequency pool, with equal proportions of all frequencies, disallowing repetitions of a frequency on successive tone pips, in order to minimise the opportunity for adaptation to frequency.

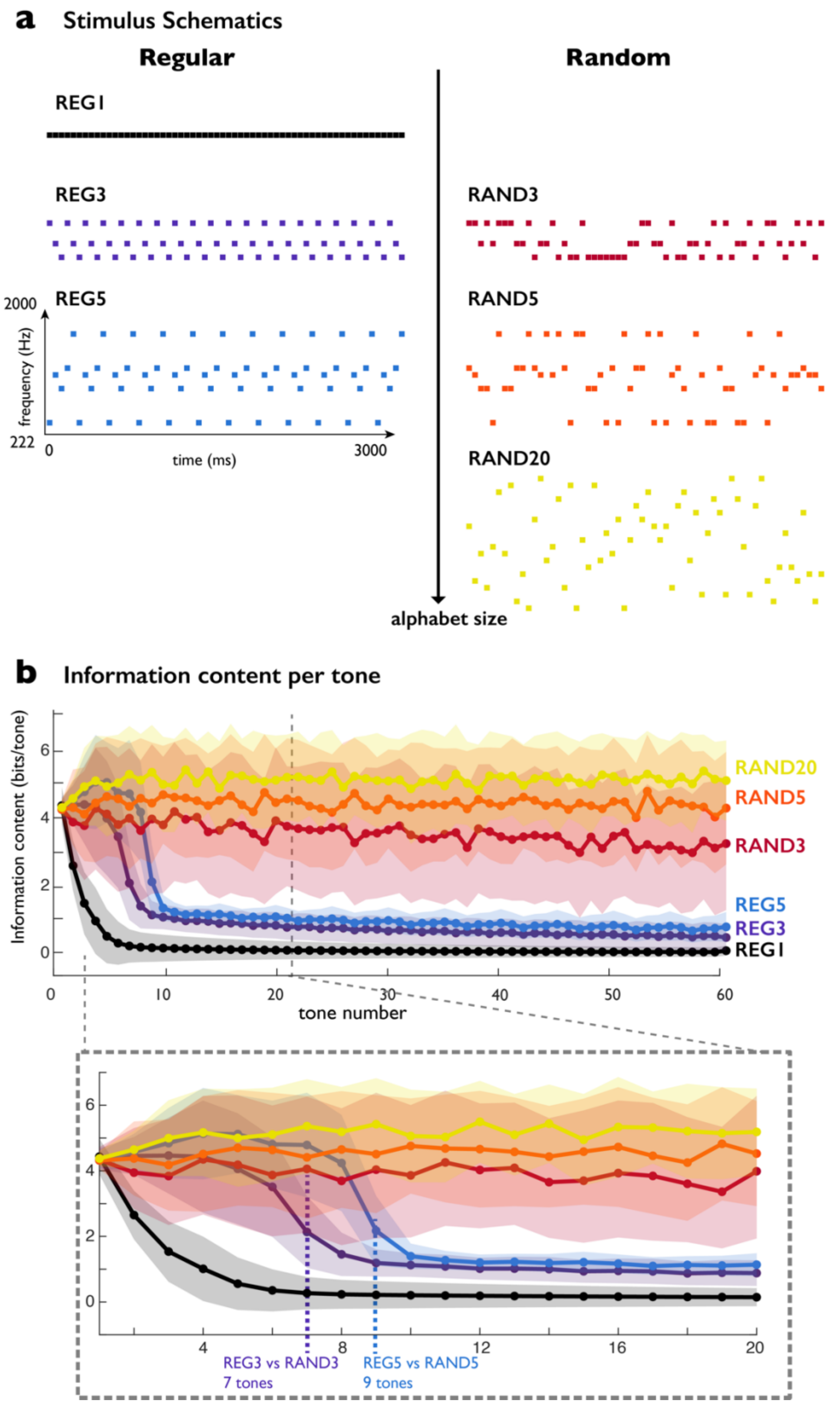

Figure 4.1 Stimuli

Tone-pip sequences varied in alphabet size alph from 1 to 20, and were either regular (REG) or random (RAND). **a:** Schematics depicting typical frequency patterns for the six conditions; squares represent individual tone-pips of duration 50 ms. For alph = 3 and 5, REG and RAND were generated in matched pairs with the same frequency content. **b:** Mean information content per tone, estimated from a model of expectation computed over an entire experiment session. Mean information content computed over 100 trials of each condition. Shaded regions around lines represent standard deviation over trials. Bottom plot shows the same data re-plotted with a zoomed-in scale to show divergence between information content at the start of the sequence as regular pattern emerges.

4.4.2 Procedure

Trials were presented in random order and each trial consisted of the presentation of a unique sequence of tone-pips, generated according to one of the 6 conditions described above. The inter-stimulus interval (ISI) was jittered between 1100 and 1500 ms. Each condition was repeated 108 times over the entire experiment, for a total of 648 trials overall. Equal proportions of each condition were presented in each of 6 blocks, which lasted for approximately 10 minutes including self-timed breaks.

The experiment was conducted in a soundproof booth while subjects watched a muted, subtitled film of their choice. Stimuli were presented binaurally using the Psychophysics Toolbox (Kleiner et al. 2007) for Matlab, using insert earphones.

4.4.3 Ideal observer model

The different conditions were modelled using an ideal observer model (Pearce 2005; Pearce et al. 2010; Barascud et al. 2016). This model tracks the transition probabilities between successive tones. From the model an estimate of the information content of each tone was derived, which can be interpreted as the level of surprise associated with each tone; the lower the value the more predictable it is. This model has previously been shown to closely correspond to the sustained response power and to the times at which brain responses diverge from one another (Barascud et al. 2016). IDyOM was used to simulate the detection of sequence predictability for 648 trials in random order, which was equivalent to the order of a true experimental session. The LTM+ model configuration was used (see Chapter 2 for more details), which initiates with an empty model, then learns over the stimulus set, updating the model after each tone. The information content of each tone was averaged over trials of the same condition, and the standard error estimated with bootstrap resampling with 1000 iterations. Pairs of conditions were compared statistically using a FWER-correction procedure with cluster-wise correction (Maris and Oostenveld 2007) at an alpha level of p<.05. The resultant plots are provided alongside the stimulus schematics in Figure 4.1, showing the dynamics of the surprise associated with each tone as regularities become established.

All conditions start with the same information content, then start to diverge. The information content for REG1 drops substantially from only the second tone, indicating that the model rapidly comes to 'expect' tone repetition. REG3 and REG5 show a significant, steep drop in information content at 7 and 9 tones respectively, suggesting that it takes a cycle plus four additional tones for the model to learn the regularity. However, during tones 2-4 for *alph* = 3, and for tones 3-6 for *alph* = 5, the reverse effect is actually seen: the information content of RAND is slightly lower than that of REG. This is possibly due to the fact that adjacent repetitions of the same frequency are possible in RAND, but not in REG. As the model learns a strong prior probability for repeats from REG1, any adjacent repeats occurring in RAND (with probability 1 in 3 or 1 in 5 for *alph* = 3 and *alph* = 5 respectively) would lead to a lower information content than different successive tones. The information content for REG1 levels off at a value lower than all the conditions, suggesting REG1 is indeed more 'predictable' than even the other REG conditions.

RAND3, RAND5 and RAND20 show fairly stable information content throughout the sequence, but the information content is significantly lower for successively smaller alphabets, indicating the model successfully detects the stochastic predictability due to restricted alphabet size.

4.4.4 Recording and data preprocessing

EEG signals were recorded using a 64-electrode Biosemi system at a sampling rate of 2048 Hz. Data were analysed using the Fieldtrip toolbox (www.fieldtriptoolbox.org/) (Oostenveld et al. 2010) for Matlab (2017a, MathWorks). Data were split into 5000-ms epochs, with 1000 ms before stimulus onset and after offset. An anti-aliasing 100 Hz low pass filter was applied before down-sampling at 256 Hz. Detrending was applied by fitting a linear trend over 9-second segments, taken from the raw recording, centred on each epoch (de Cheveigné and Arzounian 2018; see also §2.3.4).

Outlier channels and trials were removed manually using Fieldtrip's visual artefact rejection tool (see Chapter 2 for details). Around 95% of trials were retained. Further signal-to-noise improvement was achieved using DSS (de Cheveigné and Parra 2014) to maximise the reproducibility over all trials, keeping the two components which explained the greatest proportion of the variance over trials. Finally, data were re-referenced to the average over all channels.

Three phases of the evoked response were analysed, the first two represent the sequence-evoked response, and the third the offset-evoked response. (i) The onset-evoked response was considered to be from 0-300 ms post-onset, including the highly-stereotyped N1 and P2 waves of the auditory-evoked response. (ii) The sustained response, from 300-3000 ms, where the response power over all channels increases, resembling a DC-shift: polarity reversals generally cease, and this is where the effects of complex sequence regularity are expected to begin. (iii) The offset-evoked response, which resembles the N1-P2 complex, but has been shown to be influenced by the statistics of the preceding sequence (Andreou et al. 2015). For ERP plotting and statistical analysis, all evoked responses were low-pass filtered at 30 Hz.

4.4.5 Sequence-evoked responses

Epochs were baseline-corrected relative to the interval 200ms before sequence onset. Epochs were averaged over trials for each condition, yielding the evoked response in each channel. However, for statistical analysis, this evoked response was quantified as the root-mean-square (RMS) over channels, giving a measure of overall evoked power at each time sample; as in Chapter 3.

Statistical analysis was performed separately for the onset and sustained portions of the sequence-evoked response. A two-tail T-test was used to find contiguous clusters of time-points showing pairwise differences between conditions, with the threshold alpha level set at 0.05 (0.025 per tail). Correction for multiple comparisons over time used a non-parametric approach to estimate the significance of the effect in each cluster, using as the test statistic the summed t-values within each cluster. To estimate the null distribution of this statistic, a Monte Carlo permutation approach with 10000 iterations was used (Maris and Oostenveld 2007). This measure renders transient differences less likely to cross the threshold than temporally-extended ones. The existence of sustained differences between conditions will increase the threshold for significance and thus render it less sensitive to any additional effects which occur over short intervals resulting in a smaller summed t-statistic (Maris and Oostenveld 2007). For this reason, a separate a-priori latency range from 0 ms to 300 ms after stimulus onset was chosen to analyse the onset-evoked response. The sustained response was analysed in the time range 300-3000 ms. Subsequently, bootstrap resampling was used to estimate the mean and the standard error of the responses over subjects, for plotting of the grand averages shown in Figure 4.2a-e.

The following pairs of conditions were contrasted: REG5 vs RAND5; REG3 vs RAND3, REG5 vs REG1; REG5vs REG3; REG3 vs REG1; RAND3 vs RAND5; RAND3 vs RAND20; RAND5 vs RAND20; REG1 vs RAND20. These contrasts correspond to the hypotheses presented in §4.3.2.

4.4.6 Offset-evoked responses

The sequence-evoked response displays a prominent peak at offset in the RMS plot (Figure 4.2), which was analysed separately as follows. The sequence-evoked epochs were high-pass filtered at 2 Hz and baseline-corrected in the 100-ms window preceding stimulus offset, in order to remove any differences in the slow, sequence-evoked response between conditions. These data were then trimmed to epochs with 100 ms before and 300 ms after stimulus offset.

The offset-evoked response was analysed over a subset of channels showing the same response polarity. Taking an RMS over all channels would misrepresent the shape of such a response, as it contains zero-crossings owing to the baseline-correction directly preceding offset. The average signal was taken over the 10 channels showing the strongest negative deflection of the offset response over all subjects and conditions (from 50-150ms post-offset, in order to capture the initial N1-like deflection). This fronto-central group of channels comprised AFz, AF3, Fz, F1, F2, F3, F4, FCz, FC1, FC2; see topography in Figure 4.2j. The resulting timeseries (one per subject and condition) were compared pairwise, from 3000 ms to 3300 ms, utilising the same cluster-based permutation procedure as for the sequence-evoked response. The latency of the N1 offset peak was computed, for each subject and condition, as the time of the most negative value between 50 and 150 ms post-offset. This was entered into a one-way ANOVA with a factor for each condition.

4.4.7 Tone-locked responses

Whilst the responses to individual tone pips are overlapping, it is possible to estimate the strength of the onset responses to each tone pip, by calculating response power at 20 Hz, corresponding to the ITI of 50ms. The power at 20Hz was computed in individual trials between 1000 and 3000 ms; a window chosen to represent the sequence-evoked response once it has stabilised. Frequency resolution for a discrete Fourier transform depends on the length of the input data; the difference between adjacent frequency bins in Hz is given by $\frac{1}{length\ (s)}$. To obtain a high frequency resolution, these 2-second epochs were taken from individual trials, then concatenated to form a single ‘epoch’ per subject and condition. In order to compensate for different numbers of trials for each subject owing to outlier removal, the concatenated trial was trimmed to 51200 samples for each subject for each condition, to give a frequency resolution of $\frac{1}{200}Hz$. This approach isolates activity with consistent phase across trials, because trials were concatenated consistently at the onset of the 20th tone in the sequence. Power over the entire frequency spectrum was calculated using the fast Fourier transform implemented in Fieldtrip’s ft_freqanalysis function, using the multitaper method

(Thomson 1982). The resulting power spectra were then RMS-averaged over all channels. The power at 20 Hz was then normalised relative to power of surrounding frequencies, to give a signal-to-noise ratio (SNR). The 'signal' was computed as the RMS over the five bins centered on 20Hz. The RMS over the remaining bins between 19 and 21 Hz were taken as 'noise'; and the SNR computed as $\mathrm{SNR} = \frac{\mathrm{signal}}{\mathrm{noise}}$. The effect of condition on 20Hz SNR was assessed using two statistical tests. First, a 1-way ANOVA with the 6 conditions as 6 factors. Second, a 2-by-2 factorial ANOVA with one factor for *alph* (3 or 5) and another for regularity (REG or RAND) was run on the 20Hz SNR from the four relevant conditions.

4.4.8 Cycle-rate responses

It is possible that there is also periodicity in the brain response corresponding to the duration of the repeating cycles in REG3 and 5. Because these sequences were unique on each trial, this effect would most likely be manifest in the induced activity, i.e. non-phase-locked oscillation. However, the inter-trial phase coherence (ITPC) at the cycle rate was also computed. The frequencies of interest correspond to the cycle durations of REG3 (150 ms) and REG5 (250 ms); 6.67 and 4 Hz respectively. The fast Fourier transform was computed for each trial using the multitaper method (Thomson 1982), for frequencies between 1 and 8 Hz. The entire trial was used, in order to gain as high as possible frequency resolution, and in addition the trial duration was zero-padded to bring the number of samples the next power of 2, corresponding to 8 seconds to give a frequency resolution of $\frac{1}{8}$ Hz. For the power spectral density, the SNR at the frequencies-of-interest was computed as for the tone-locked response, but as the frequency resolution was much lower, only a single bin was taken the signal and the surrounding 2 bins on each side as noise. To compute the ITPC for each subject and condition, the complex vectors representing phase and power at the frequency-of-interest in each trial were normalised to unit length, then averaged. The modulus of this averaged vector represents the consistency in phase across trials with a value between 0 and 1 (Lachaux et al. 1999), and the angle of the vector represents the predominant phase. A paired-sample t-test was used to test for differences between REG and RAND, in response power and ITPC, at the cycle rate for *alph* = 3 and *alph* = 5.

4.4.9 Participants

21 normal-hearing young participants took part, with no known neurological or hearing problems. One subject was excluded due to noise during the EEG recording, leaving 20 subjects, 6 male, aged between 20 and 33 (mean 24.4). None of the subjects had participated in similar experiments conducted in the lab.

4.5 Results

I investigated effects of sequence regularity and alphabet size on evoked responses by contrasting the power of the response between different conditions over peristimulus time. I found effects of regularity and alphabet size throughout the sustained response, as well as more transiently at stimulus onset and offset.

4.5.1 Sequence-evoked response

Figure 4.2a-e shows the average voltage (RMS over channels) of the evoked response. For clarity, only a subset of conditions is shown in each plot. Comparisons were made between certain pairs of conditions for which a hypothesis was described in §4.3.2. Statistical analysis employed a permutation procedure to correct for multiple comparisons over time (§4.4.5), to derive the following significant temporal clusters at FWER< 0.05; shown in Figure 4.2 as horizontal bars below the evoked responses.

All conditions are characterized by a series of onset peaks corresponding to the N1 and P2 time windows, a rise to a plateau (sustained response) of around 2µV for the remainder of the trial, and in some cases a clear offset peak. The offset response is discussed separately below.

4.5.1.1 Regularity increases the sequence-evoked response for *alph* = 3 and 5

The effect of sequence regularity, controlled for (long term) tone probability, is seen when contrasting REG and RAND of the same alphabet size (Figure 4.2a-b). Based on previous EEG (Southwell et al. 2017) and MEG (Barascud et al. 2016) findings with similar stimuli, it was predicted that the sustained response to REG would be greater than to RAND throughout the entire sequence, both for alph = 3 and 5. Indeed the sustained response is greater for REG5 than for RAND5 throughout the sequence. REG5 diverges from RAND5 at 453 ms, which is immediately following the fourth tone in the first repeated cycle; exactly as predicted from the ideal observer model (see Figure 4.1b). The REG5 sustained response remains significantly higher in five clusters covering most of peristimulus time. From the ideal observer model, the divergence between REG and RAND was expected to be earlier for REG3 than REG5, at around 350 ms which corresponds to 7 tones. The difference between REG3 and RAND3 is significant from 359 ms, which closely matches the prediction. However REG3 does not entirely behave as hypothesised; whilst the sustained response is initially higher for REG3, from 1051 ms it dropped back to a similar level as RAND3, remaining statistically indistinguishable for the remaining 2 seconds of the stimulus. The response to REG3 initially rises above REG5, likely reflecting the earlier discovery of regularity, rising significantly above REG5 from 258-551ms (Figure 4.2c).

4.5.1.2 RAND response power is inversely related to *alph*

The effect of alphabet size, independent of regular cycle length, is revealed by the comparison between RAND3, RAND5 and RAND20 (Figure 4.2d). This represents a manipulation of non-deterministic predictability from the most predictable (RAND3) to the least (RAND20). The response to RAND3 is significantly greater than RAND20, from 188-230 ms after stimulus onset, then again from 520 ms until the end of the sustained response. Likewise, the response to RAND5 is higher than to RAND20, but significantly so for less of the sequence (between 527 and 2137 ms). There was no significant difference between RAND3 and RAND5; although the mean sustained response level appears to be consistently higher for RAND3 than RAND5, this is a small difference.

4.5.1.3 REG1 responses do not behave as predicted

In order to investigate the effect of simple tone repetition in regular conditions, REG1 was compared to REG3 and REG5 (Figure 4.2c). In the onset response, REG3 showed a briefly larger peak than REG1 in the range 180-215 ms. Both REG3 and REG5 evoked a larger sustained response than REG1, despite REG1 containing a lower information content than the other REGs (according to the IDyOM modelling). This first reached significance at 625 ms for REG3 and 879 ms for REG5. Both contrasts with REG1 did not show stable significance throughout the sequence, although significant clusters were seen for REG5>REG1 until offset. Response power to REG3 dropped throughout the sequence and was no longer significantly different from REG1 after 2164 ms; although the grand average power never drops to that of REG1.

REG1 was compared to RAND20 (Figure 4.2e), the least repetitive stimulus, to determine if repetition suppression or repetition enhancement is dominant in the sequence-evoked response. At onset, repetition suppression was seen with RAND20 significantly above REG1 from 176-227 ms. During the sustained response the opposite effect was seen; REG1 greater than RAND20 from 402 ms, continuing to be a significant effect intermittently throughout the trial.

To summarise; REG sequences evoked a higher sustained response than matched RAND sequences, for alphabet size 3 and 5, although this was only the case for the first second of the sequence for REG3. Within random sequences, decreasing *alph* was associated with an increased sustained response, replicating Chapter 3. REG1 did not behave like the other five conditions, exhibiting a nuanced combination of repetition suppression and enhancement. REG1 evoked a consistently lower response throughout the trial than REG3 or REG5.

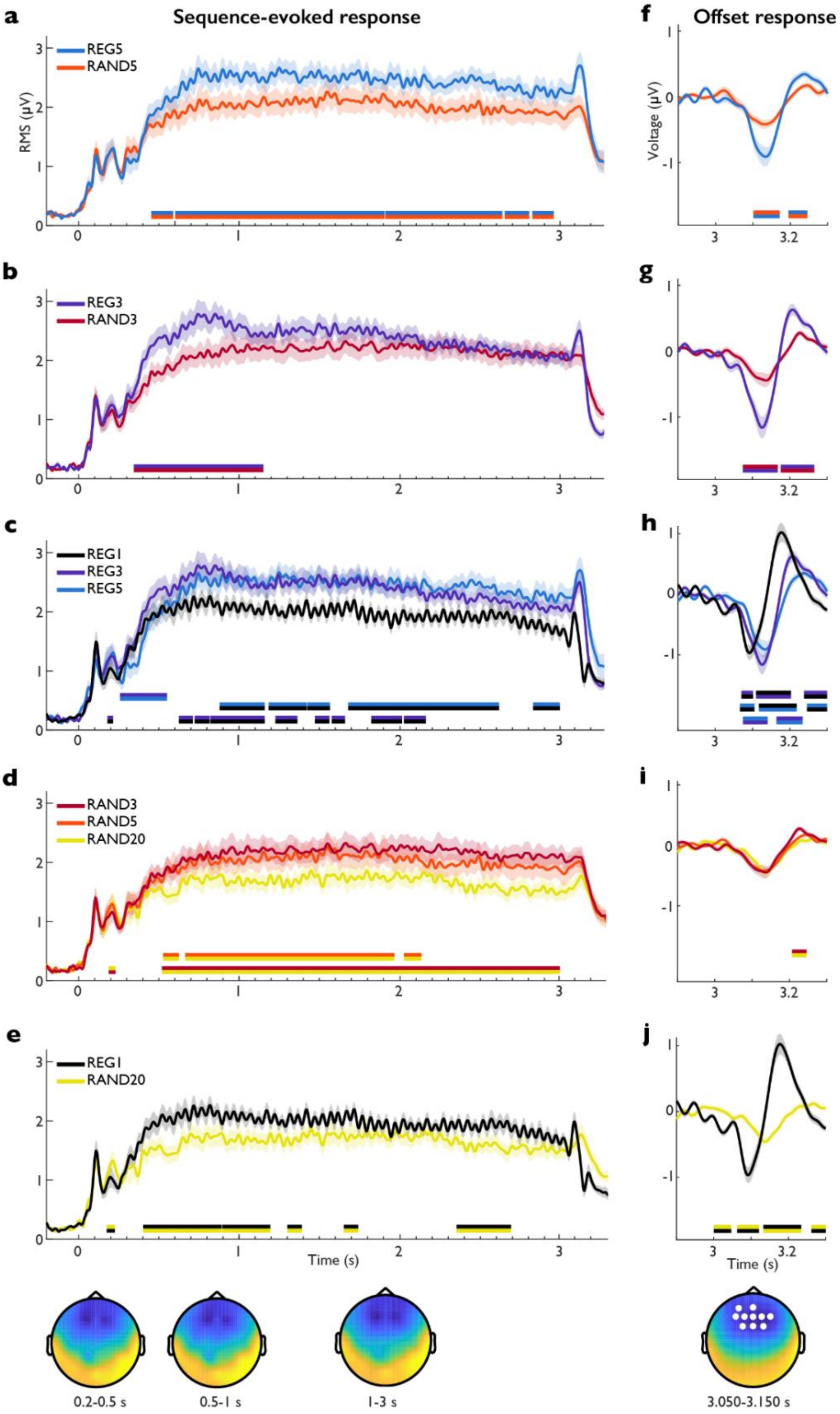

Figure 4.2 Evoked responses

Bars under plots denote periods of significant difference between conditions; the two colours indicate which pair of conditions are represented. Sustained response shown in a-e; offset response shown in f-j. The lower panel shows topography averaged over all conditions during the specified time ranges. The topography below j indicates 10 channels over which the offset-evoked response is averaged in white.

4.5.2 Offset-evoked EEG response

The offset responses were analysed separately, removing the pre-existing sustained response differences through baseline correction and high-pass filtering (see §4.4.6). These evoked responses, averaged over a selection of 10 channels to retain the polarity reversals of successive peaks, are shown in Figure 4.2f-j. All responses resemble the classic N1-P2 responses to sound onset, so are referred to as offset N1 and offset P2 below. One-way ANOVAs were carried out on the peak latency and peak magnitude of the offset N1, for the peaks identified separately for each subject and condition. There was a main effect of condition on peak magnitude ($F_{5,119} = 7.3$, $p < 0.0001$) and latency (one-way ANOVA; $F_{5,119} = 7.7$, $p < 0.0001$). Post-hoc comparisons used Tukey's honestly significant difference procedure and are reported below.

4.5.2.1 Regularity increased the offset-evoked response for *alph* = 3 and 5

Figure 4.2 f-g shows the offset response for REG was substantially larger than for RAND, for *alph* = 3 and 5. This effect was significant for the offset N1 from 74-168 ms for *alph* = 3; and from 102-172 ms for *alph* = 5. For the offset P2, REG was more positive than RAND from 195-245 ms for *alph* = 5 and from 176-266 ms for *alph* = 3. REG3 showed a greater negative deflection than REG5 from 74-141 ms, and a correspondingly larger P2 than REG5 from 164-234 ms. These significant effects for REG3 as compared to RAND3 and REG5 are interesting, as the sustained response for REG3 was not different directly preceding offset, suggesting that the offset and sustained responses probe different processes.

4.5.2.2 The offset N1 occurred earlier for REG1

Figure 4.2h shows offset responses to the three REG conditions. The offset N1 falls significantly earlier than for the other conditions, with a mean latency of 96 ms, as opposed to over 116 ms for all other conditions. The latency of the N1 did not differ for all other conditions. The N1-latency response appears to be smaller than for REG3 and REG5, however according to post-hoc comparisons the REG1 offset N1 was not significantly different from either REG3 or REG5. REG1 again shows an earlier peak in the P2 window, but it is somewhat larger than for either of the other conditions. As this effect is likely correlated with the earlier N1 latency, the P2 peak was not subject to a separate statistical analysis.

The comparison between REG1 and RAND20, respectively the most repetitive sequence and the least, resulted in significant differences throughout the offset epoch, beginning at 0 ms. This implies that ongoing differences in the sustained response between REG1 and RAND20 have not adequately been removed, as it is not possible for an auditory evoked response to occur with such a short latency. This might reflect the increased tone-locked response for REG1 during the final few tones of the sequence, which leads to larger, transient 'troughs' in the evoked response just before offset (see §4.5.3); however this seems to be superimposed on a slower downward trend; which could reflect adaptation in the slow response to REG1. Alternatively it is possible that the acausal high-pass filter introduced spurious differences between REG1 and RAND20 before they really occur (see de Cheveigné and Nelken 2019, and §2.3.4.1). Nevertheless, it does appear that the offset responses are much larger and more pronounced than to RAND20, confirmed by the ANOVA on offset N1 magnitude.

4.5.2.3 The effect of *alph* on the RAND response

All RAND conditions show the same offset response profile, but greatly attenuated in magnitude as compared to the REG conditions. The only difference is seen between RAND3 and RAND20 after 207 ms, with RAND3 showing a slightly larger P2 response.

Overall, the offset response effects mirror those seen in the sustained response, reflected instead in a phasic, brief response rather than a tonic, temporally extended potential. However, notably, there is a more pronounced offset response for REG3 than RAND3 despite the sustained responses being statistically indistinguishable for the latter two seconds, and the means overlapping. Moreover, the offset P2 to REG1 is larger than for REG3 and REG5 despite exhibiting a lower sustained response.

4.5.3 Tone-locked response

The power spectral density (PSD) around 20Hz for all conditions is shown in Figure 4.3a. It can be seen that there is a highly pronounced peak at 20Hz for all conditions. The signal-to-noise ratio (SNR) of the response at 20Hz is shown in Figure 4.3b, representing the tone-locked response; this is much larger for REG1 than for all other conditions. A one-way repeated-measures ANOVA revealed an effect of condition on the tone-locked response ($F_{5,119} = 13.35$, $p < 0.0001$). Post-hoc comparisons were performed using Tukey's honestly significant difference procedure (Tukey 1949), revealing a single significant difference between REG1 and all other conditions. As the four conditions REG3, REG5, RAND3 and RAND5 represent a balanced design, the effects of regularity and of alphabet size on the tone-locked response were analysed using a two-way repeated-measures ANOVA, with factors of regularity and *alph* (Figure 4.3). This shows a main effect of regularity ($F_{1,19} = 5.48$, p = 0.030), a main effect of alphabet size ($F_{1,19} = 4.69$, $p = 0.043$) and no interaction ($F_{1,19} = 1.636$, $p = 0.22$); the SNR at 20Hz being higher for REG than for RAND, and higher for *alph* = 3 than *alph* = 5. That there are effects of regularity and alph for the 2-way ANOVA but no differences between any of the comprising conditions, indicates that the effect of regularity is too small to detect in individual conditions.

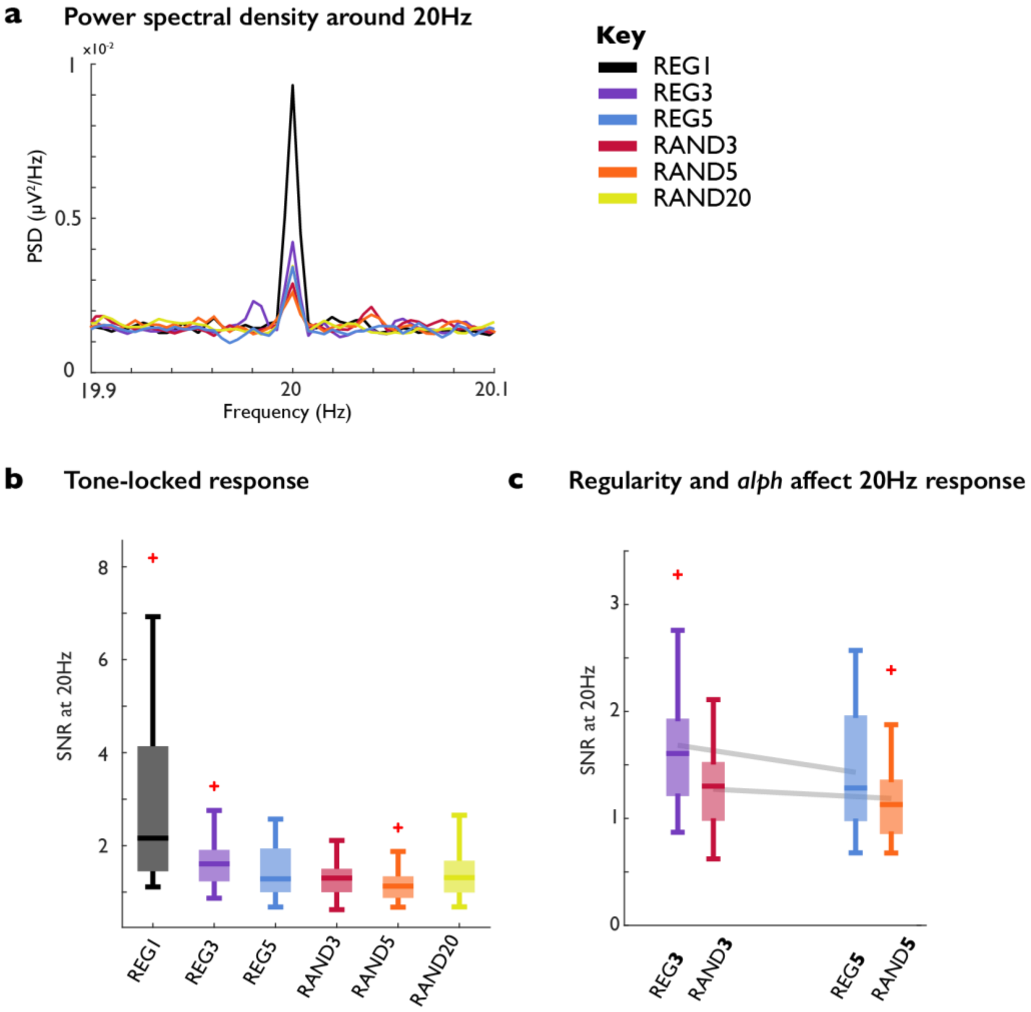

Figure 4.3 Tone-locked response

Frequency analyses. **a:** Power spectral density around 20Hz. **b:** Signal-to-noise ratio at 20Hz for each condition. **c:** Signal-to-noise ratio at 20Hz for matched REG and RAND conditions. Grey lines connect group means for REG and RAND of different *alph*. Boxplots show distribution over subjects: median with lower and upper quartiles, whiskers show range. Extreme observations, falling outside of 1.5 times the interquartile range from the start of the whisker, are shown by a red cross.

4.5.4 Cycle-rate response

There is a significant effect of regularity on the cycle-locked response at the cycle rate in both REG3 and REG5 as compared to their matched RANDs (Figure 4.4 a&b). REG3 shows a larger response at 6.67 Hz than RAND3 (t_{19} = 3.03, *p* = 0.0068), and REG5 shows a larger response at 4 Hz than RAND5 (t_{19} = 2.40, *p* = 0.0268). This means that the periodicity of REG is reflected in the response on individual trials, but this is likely to be an idiosyncratic pattern which does not have a consistent phase across trials, because the regularity comprises a different pattern each time. Indeed, the ITPC did not significantly differ between REG and RAND (*alph* = 3: *p* = .62 , *alph* = 5: *p* = 0.51; Figure 4.4 c&d).

A similar effect was seen by Herrmann and Johnsrude (2018), who measured EEG oscillatory responses to RAND and REG stimuli similar to the present chapter, finding no difference in ITPC between REG and RAND; as was the case here. The authors also included a REG condition where the tone frequencies themselves followed a sinusoidally frequency-modulated (FM) pattern with consistent phase across trials. They found the ITPC to be greater in REG which followed such an FM pattern, than in matched RAND sequences. However, they did not look at a phase-insensitive measure of the cycle-rate response such as the induced response power as used here.

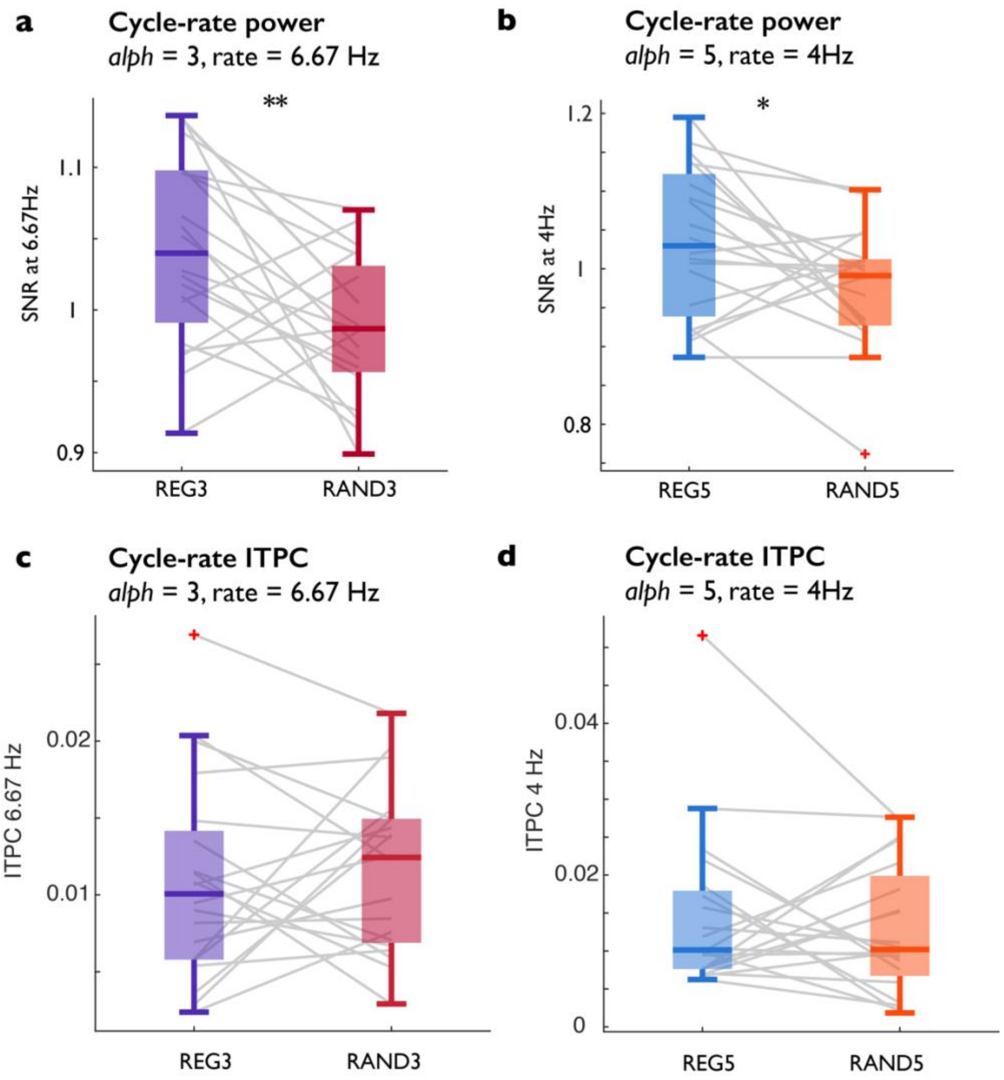

Figure 4.4 Cycle-rate response

a: Signal-to-noise ratio at cycle rate of 6.67Hz for *alph* = 3. **b:** Signal-to-noise ratio at cycle rate of 4Hz for *alph* = 5. **c:** ITPC at 6.67Hz for *alph* = 3. **d:** ITPC at 4Hz for *alph* = 5. Boxplots show distribution over subjects: median with lower and upper quartiles, whiskers show range. Extreme observations, falling outside of 1.5 times the interquartile range from the start of the whisker, are shown by a red cross. . $^{**}p < .01$, $^{*}p < .05$

4.6 Discussion

4.6.1 Summary of results

To summarise, predictability in two forms is associated with an increase in brain response; whether the predictability arises from a deterministic order of frequencies being repeated (REG) or from a restricted alphabet size (RAND3 as compared to RAND5 and RAND20). However this was not universally the case; notably the response to REG1 behaved differently from the rest in several respects. The sustained response to REG1 never reached as high as for REG3 and REG5, despite being equally predictable. Yet, the response to maximally-repetitive REG1 was still higher than for the least adapting stimulus, RAND20. This suggests that opposing effects of repetition suppression and regularity enhancement influence the sustained response.

4.6.2 Onset effects

The long sequences offer the opportunity to observe how the brain response unfolds over several seconds. The manipulations in this experiment affected different phases of the evoked response in different ways. A reduced response was seen in the first 300 ms for REG1 relative to other conditions with a larger alphabet, and for RAND3 relative to RAND20. This could be explained as adaptation to tone frequency, as in all these cases the suppressed condition contains more repetitions over frequencies encountered in the sequence so far. The sub-300-ms timescale of these effects is in line with the timing of repetition suppression observed in multiple paradigms (Budd et al. 1998; e.g. Grill-Spector et al. 2006; Herrmann et al. 2015). This onset effect precedes the rise to the sustained response when, it is argued in this thesis, the representation of predictability influences the response magnitude.

4.6.3 The sustained response reflects predictability for complex sequences

For the sustained response, some of the hypothesised effects were indeed seen. The more restricted the alphabet in random sequences, the greater the evoked response magnitude; which is in line with the decreased information content of each tone in RAND with a restricted alphabet (Figure 4.1b). Although this is surprising from the standpoint of neuronal adaptation, there is accumulating evidence that the sustained response encodes this aspect of sequence predictability. Similarly, regularity boosted the sustained response over random sequences of the same tone alphabet. The timing of the initial divergence between REG and RAND was related to alphabet size as predicted. However, for REG3 this effect was transient before the evoked power dropped back to the level of RAND3, following around 7 repetition cycles. This decrease could be explained by expectation suppression: as the regularity is extracted, top-down expectation signals serve to suppress evoked responses. However, this explanation is at odds with the account that the sustained response represents the level of predictability in the sequence. It is entirely possible that the response to REG stimuli with larger alphabets would also exhibit a similar drop in power eventually, but the sequence would need to be longer for this effect to be seen. Likewise the 20 Hz response was enhanced by regularity, for *alph* = 3 and 5, though this effect was small enough that it is not significant when individual conditions are analysed.

4.6.4 Offset responses reflect sequence predictability

A different way of probing the brain's internal model of stimulus regularity is to violate the pattern and observe the deviance response. This will be covered in detail in Chapter 5, but in the current experiment there is a form of deviance consisting of the stimulus offset (Andreou et al. 2015); this violates the simple expectation of each tone pip being followed by another, established over the preceding three seconds. Indeed, when the sustained response is subtracted away and epochs analysed following offset, an N1-P2 series is observed (Picton et al. 1978). The offset response to REG was larger than for alphabet-matched RAND. Interestingly, the offset response to REG3 was markedly larger than to RAND3, despite the sustained response showing no difference immediately preceding offset. The offset-evoked response can be considered a precision-weighted prediction error, where the error in question refers to the violated expectation that a sound will occur. The precision-weighting would be largest for REG, then RAND in increasing order of *alph*.

An alternative explanation for the existence of the offset response is that it reflects adaptation to the rhythmic stimulus stream. May et al. (2001) hypothesised the existence of neural circuits in auditory cortex implementing selectivity for particular auditory modulation rates, consisting of coupled excitatory and inhibitory neurons, which have a particular resonant oscillatory frequency. This model entailed the prediction that these micro-circuits would continue to oscillate temporarily on cessation of an isochronous stimulus train, producing evoked responses resembling the response to the sounds themselves. With empirical MEG data they showed that, at least in the case of isochronously-presented tones, there is such a rebound response at offset. This model could explain the REG versus RAND effect in the offset response here; as REG can be considered as several interleaved isochronous streams of tone pips at a given frequency. This model was formulated for, and tested on, tones at a single frequency, and is not explicit whether the rate-sensitive circuits are also frequency-selective. If frequency- and rate- selective circuits exist, they would entrain to the regular occurrence of each individual frequency in REG, then lead to an offset response as the oscillations continue. However this exact model cannot fully account for the present results. Both the model and the empirical results showed a linear relationship between the stimulus presentation rate and the latency of the offset responses; such a relationship does not seem to be present here as the offset latencies are very similar across alphabet sizes. Also, the model of May et al. postulates that the rate-selective circuits are active during stimulus presentation, showing an even stronger oscillation than that present after offset. The present sustained responses show periodic activity at a much lower amplitude than the observed offset responses.

4.6.5 REG1 does not behave as predicted

Notably, the REG1-evoked response was lower than for REG3 and REG5, despite being as predictable, which strongly indicates that adaptation at least has a sizeable contribution to the observed sustained responses. REG1 represents the most predictable stimulus, which also theoretically leads to the greatest degree of adaptation in units selective for the repeated frequency. Here, repetition suppression is indeed seen. Whether the mechanism is adaptation, or whether it also depends on prediction-generated suppression from higher-level areas remains unknown. What is clear is that the responses to this stimulus set cannot solely be described by repetition suppression, as demonstrated most strongly by the comparison with RAND20. The response was still higher than to RAND20, both during the sustained response and at offset. Adaptation caused by neuronal refractoriness may be ongoing throughout the sustained response for more complex sequences, but is mostly offset by other processes.

Tone-locked responses at 20 Hz for REG1 were considerably enhanced relative to all other conditions. This 20-Hz response may reflect the summed result of overlapping onset responses to tone onsets. Although an adaptation-based account would predict tone-onset response to be lower for REG1 due to low-level adaptation. the opposite effect was seen quite strongly. This would be consistent with increased gain due to the precise expectation of the repeated frequency. However, to speak against the precision-based account, the 20Hz response did not appear to be affected by regularity as strongly for more complex sequences consisting of multiple frequencies. Alternatively, the increased 20Hz response could instead be a reflection of more precisely time-locked responses to each tone. As the frequency is the same for each tone, the same neural population responds each time, perhaps manifesting as less temporal jitter in the scalp-level response. Regardless of the mechanistic explanation, it is clear from the 20Hz response that the REG1 response exhibits very different behaviour than for all other conditions. The offset response to REG1 is also earlier than for all other conditions, but not larger in magnitude. This could again be due to more precisely time-locked tone-onset responses than for other conditions, such as the earlier offset response seen when manipulating temporal regularity (Andreou et al. 2015). These findings provide evidence of opposing processes underlying the influence of tone repetition on the evoked response. To put this in a wider context, the differences between the REG1 response as compared to all other conditions indicates tone repetition does not form a generalisable model of regularity.

4.6.6 Repetition suppression or repetition enhancement?

Overall, these results reveal opposing effects underlying the sustained responses to rapidly-presented sequences. The dichotomy of repetition suppression and repetition enhancement are exhibited for REG1 over different timescales. This can be related to examples of overlapping yet dissociable repetition-evoked phenomena in the literature (Todorovic and de Lange 2012; Grotheer and Kovács 2015).

4.6.6.1 Dissociable repetition-related responses

Predictability-related suppression of measured responses can result from a combination of different underlying processes, whether neuronal refractoriness, prediction error suppression or sharpening of representations (Desimone 1996; Budd et al. 1998; Grill-Spector et al. 2006; Kok, Jehee, et al. 2012). Even at the neuronal level within a given cortical area, adaptation to repetitive stimuli affects excitatory and inhibitory inputs at different timescales (Solomon and Kohn 2014; Chen et al. 2015), and has complex downstream effects (Solomon and Kohn 2014). At the population level, different timescales of repetition effects correspond to different cognitive phenomena. Expectation suppression and repetition suppression to sound have been demonstrated to have distinct temporal profiles (Todorovic and de Lange 2012), with repetition affecting earlier response components (between 40 and 60 ms) than expectation (from 100 to 200 ms). This effect was generalised to the visual modality in fMRI (Grotheer and Kovács 2015) and EEG (Feuerriegel et al. 2017). These findings support a two-stage account of repetition suppression, incorporating a hierarchically-higher stage of expectation suppression following an earlier primary-sensory stage of repetition suppression (Grotheer and Kovács 2016). The divergent effects on the sustained response of simple repetition (REG1) and complex regularities (the other 5 conditions) seen here supports a similar model.

These studies reveal that stimulus structure can influence different hierarchical levels of processing in different ways, even in early evoked responses. Although the present experiment involved passive listening in naïve subjects, so any expectations formed on the basis of regularities are most likely implicit, a similar multi-stage model could account for at least some of the results here. However, distinct from the cited findings, this study revealed multiple instances of enhancement according to predictability, even for the simplest model of predictability engendered by the repetition in REG1. Repetition enhancement, although less investigated than repetition suppression, has been reported in multiple domains (Segaert et al. 2013). It is most often observed in explicit paradigms requiring behavioural responses such as priming (e.g. Petit et al. 2006; Henson et al. 2008) but it has been reported in responses to passively-presented stimulation (Stefanics et al. 2018). Of most relevance to the present chapter, concomitant repetition suppression and repetition enhancement effects were observed in an MEG study of passively presented auditory roving-standard sequences (Recasens et al. 2015). Repetition suppression occurred during the N1m, and was localised to temporal cortical sources. Later repetition enhancement was observed between 200 and 300 ms during a sustained potential with greater field strength than the N1m, with generators in the same temporal areas but also in inferior frontal gyrus. Although EEG and MEG have different sensitivities, it is tempting to relate this to the current EEG sustained response displaying repetition enhancement when comparing REG1 and RAND20.

4.6.6.2 Predictive coding interpretation

Predictive coding provides two neuronal mechanisms for response enhancement (precision and prediction), and one mechanism for response suppression (reduced prediction error), consistent with the processing of a predictable stimulus stream. These accounts apply not only to immediate, exact repetition (as in REG1) but also to more complex predictability (as in REG versus RAND, and RAND with low *alph* as opposed to high). The sequences used in this chapter represented a constant level of predictability throughout each trial, such that by the first half-second or so of each stimulus, it is possible to form an expectation of both the *identity* of the next tone and the *precision* of this estimate, given the context provided by the previous tones. The neuronal mechanism for precision signalling in predictive coding is postsynaptic gain on the superficial pyramidal prediction-error units, leading to ascending prediction errors up-weighted by (expected) precision (Friston 2005; Feldman and Friston 2010). Such increased superficial pyramidal activity plausibly forms a large contribution to measured EEG signals (Murakami & Okada 2006; see §2.3.1}. An explanation for the sustained response and tone-locked response differences in terms of increased precision remains consistent with the present results, bar the lower sustained response magnitude to REG1 as compared to REG3 and REG5.

However, the extraction of regularity may also lead to greater prediction-related activity. These expectations may be formed in higher-order areas such as prefrontal cortex, but also intrinsically within auditory cortex (Friston 2005; Auksztulewicz and Friston 2016), resulting in suppressed prediction error signalling arising directly from sensory input. However, this increased prediction activity for REG relative to RAND, which is established through descending inhibitory connections, could perhaps be detected as an increased evoked response within higher-order sensory and prefrontal areas. Response enhancement at the whole-scalp level could actually be due to activation of additional sources rather than enhancement of responses in the same neural populations, as discussed in Chapter 3.

4.6.7 Conclusion

The results presented in this chapter demonstrate that subtle interactions between enhancement and suppression of auditory responses are involved in automatic tracking of stimulus statistics, implicating multiple, parallel processes.

Chapter 5. Deviance Responses in Regular and Random Sequences

5.1 Statement of Contribution

This chapter is adapted from a published paper: **Southwell R, Chait M. 2018. Enhanced deviant responses in patterned relative to random sound sequences. Cortex. 109:92–103.**

5.2 Summary

The brain draws on knowledge of statistical structure in the environment to facilitate detection of new events. Understanding the nature of this representation is a key challenge in sensory neuroscience. Specifically, it is unknown whether real-time perception of rapidly-unfolding sensory signals is driven by a coarse or detailed representation of the proximal stimulus history. We recorded electroencephalography brain responses to frequency outliers in regularly-patterned (REG) versus random (RAND) tone-pip sequences which were generated anew on each trial. REG and RAND sequences were matched in frequency content and span, only differing in the specific order of the tone-pips. Stimuli were very rapid, limiting conscious reasoning in favour of automatic processing of regularity. Listeners were naïve and performed an incidental visual task. Outliers within REG evoked a larger response than matched outliers in RAND. These effects arose rapidly (within 80 ms) and were underpinned by distinct sources from those classically associated with frequency-based deviance detection. These findings are consistent with the notion that the brain continually maintains a detailed representation of ongoing sensory input and that this representation shapes the processing of incoming information. Predominantly auditory-cortical sources code for frequency deviance whilst frontal sources are associated with tracking more complex sequence structure.

5.3 Introduction

Detection of new events within a constantly fluctuating sensory input is a fundamental challenge to organisms in dynamic environments. Hypothesized to underlie this process is a continually-refined internal model of the real-world causes of sensations, made possible by exploiting statistical structure in the sensory input (Dayan et al. 1995; Friston and Kiebel 2009; Winkler et al. 2009; Rubin et al. 2016). Evidence from multiple domains, including speech (Saffran et al. 1996), abstract sound sequences (Saffran et al. 1999; McDermott et al. 2013), vision (Turk-Browne et al. 2009) and motor control (Bestmann et al. 2008) reveals sensitivity to environmental statistics, which in turn influences top-down, expectation-driven perceptual processing. When the organism encounters sensory input that is inconsistent with the established internal model, a 'surprise' response is generated (Friston 2005), promoting a rapid reaction to the associated environmental change. Understanding what aspects of stimuli are 'surprising', and how they are processed, is therefore central to understanding this network.

5.3.1.1 MMN reflects mismatch computation

The auditory system has been a fertile ground for probing sensory error responses, at multiple levels of the processing hierarchy (Nelken 2014; Aghamolaei et al. 2016; Ayala et al. 2016). A common approach involves using a stream of standard sounds to establish a regularity that is occasionally interrupted by 'deviant' sounds (Garrido et al. 2008; Garrido, Kilner, Stephan, et al. 2009; Khouri and Nelken 2015; Heilbron and Chait 2017). Deviants usually evoke an increased response relative to that measured for the standards (Ulanovsky et al. 2003; Garrido, Kilner, Stephan, et al. 2009; Herrmann et al. 2015). Since many of the investigated sequences have been very simple, often a repeated tone; neural adaptation is likely a major contributor to the observed deviant responses (Grill-Spector et al. 2006; Briley and Krumbholz 2013; Nelken 2014). However, accumulating evidence suggests that at least part of the deviant response arises from neural processes associated with computing 'surprise' or detecting a mismatch between expected and actual sensory input (Taaseh et al. 2011; Daikhin and Ahissar 2012; Khouri and Nelken 2015, Parras et al. 2017). The underlying network, consistently implicated in these processes, is comprised of bilateral auditory cortex (Heschl's Gyrus and superior temporal gyrus) and right inferior frontal gyrus (Opitz et al. 2002; Garrido et al. 2008; Garrido, Kilner, Stephan, et al. 2009; Barascud et al. 2016; Chennu et al. 2016; rIFG; Heilbron and Chait 2017).

5.3.1.2 Contextual influences on deviance responses

What information is used in calculating surprise? Mounting evidence suggests that the deviant response is shaped by the statistics of the sequence as it unfolds. Garrido et al. (2013) demonstrated that MEG responses to probe tones are sensitive to the statistical context (mean and variance of frequency) of randomly generated tone-pip sequences such that larger responses occurred to the same probe tone when presented in a context with low-variance than with high-variance. Rubin et al. (2016) modelled brain responses to two-tone sequences with different probabilities. They demonstrated, in line with conclusions from Garrido et al. (2013), that trial-wise neural responses in auditory cortex are well explained by the probability of occurrence of each tone frequency, calculated from the recent history of the sequence. The models that best fit neural responses were based on a relatively long stimulus history (~10 tones); but were a coarse representation, reflecting a small set of summary statistics.

Most previous work investigating the effect of context on deviant processing has focused on simple, random frequency patterns (May and Tiitinen 2010; Herrmann et al. 2015; Khouri and Nelken 2015). For these signals, a coarse representation, possibly underpinned by adaptation processes (May and Tiitinen 2010; Herrmann et al. 2015; Khouri and Nelken 2015), may indeed be sufficient to capture behaviourally-relevant attributes. However, it remains unclear whether the brain also keeps track of a detailed history of past sensory experience. To reveal these processes, the stimulus must contain some structural regularity. Whilst some previous research (Koelsch et al. 2000; Maess et al. 2001; Pearce et al. 2010; Koelsch et al. 2016), investigated complex sequence structure, the experiments mostly involved fixed patterns and exposure over very long durations, likely reflecting long-term structure learning. Here we focus on structure which emerges anew in each sequence. We seek to understand whether the brain represents this structure, and identify the underlying brain networks.

5.3.1.3 Deviance responses in REG and RAND

We used fast tone-pip sequences, unique on each trial, that occasionally contained a frequency outlier presented outside of the spectral region occupied by the standards. To determine whether the deviant response merely reflects an unexpected change in frequency between the standards and outlier, or whether it is also affected by the specific order of elements in the sequence, we used as standards either regular (REG) or random (RAND) sequences of otherwise matched frequencies (see Figure 5.1), such that the frequency span is identical but the precision of the available information regarding successive frequencies is either low (RAND) or high (REG). Notably, the sound sequences were very rapid (20 tones per second) such that conscious reasoning about the sequence order is unlikely to be possible.

Based on the hypothesis that the human brain tracks and evaluates incoming sensory information against the specific pattern established by the sequence context, we expect outlier tones to be more readily detectable in REG than in RAND sequences. The experiments reported below investigate this assertion by measuring deviance-evoked EEG responses in naïve, distracted listeners (Experiment 1) and when listeners actively monitored the sequences for outlier tones (Experiment 2).

5.4 Stimuli

Stimuli consisted of 50-ms tone pips of varying frequency, arranged in regular (REG) or random (RAND) frequency patterns over a total duration of 3000ms (60 tones).

Frequencies were drawn from a pool of 26 logarithmically-spaced values between 198 and 3563Hz (12% increase in frequency at each step; equivalent to two musical semitones). To generate each sequence, 13 adjacent frequencies were chosen at random from the larger pool (see Figure 5.1a) and then a random subset of 10 of these frequencies were retained, so that all sequences had a similar bandwidth and contained exactly 10 unique frequencies ('alphabet size' = 10). REG sequences were generated by permuting the 10 chosen frequencies and then repeating that order six times (Figure 5.1b; top). Matched RAND sequences were generated by shuffling each REG sequence, with the constraint that no two adjacent tones were the same frequency (Figure 5.1b; bottom). Overall, the stimulus generation procedure ensures that REG and RAND sequences are matched exactly in terms of the first-order distribution of tones; the only difference being whether they are arranged in a predictable (REG) or unpredictable (RAND) order.

Half of the sequences (henceforth denoted as **REG$_{\mathrm{O}}$** and **RAND$_{\mathrm{O}}$**) contained a single frequency 'outlier' tone between 1500 and 2750ms post-onset (latency chosen at random for each stimulus), which is equivalent to a minimum of 3 REG cycles. Our previous work (Barascud et al. 2016; Southwell et al. 2017) determined that the detection of regularity and the associated brain responses take place between 1-2 cycles. A latency of 3 cycles therefore assures that the processing of the regular pattern has stabilized (see also Figure 5.2a). The outlier tones replaced the corresponding standard tone. The outlier frequency was either higher or lower than the range spanned by the 10 standard frequencies in the sequence, with a minimum distance of two frequency steps. Throughout the entire set of trials, all 26 frequencies could be outliers or standards. Furthermore, to ensure all ten standard frequencies were approximately equally probable before the outlier, RAND$_{\mathrm{O}}$ were generated by shuffling separately before and after the chosen outlier position. Stimuli were generated in matched sets of four (Two containing an outlier: REG$_{\mathrm{O}}$, RAND$_{\mathrm{O}}$; and two matched 'controls' with no outlier: **REG$_{\mathrm{NO}}$**, **RAND$_{\mathrm{NO}}$**), using the same 'alphabet' for standards (and the same frequency for the outliers, if applicable). Sequences were unique on each trial and generated anew for each subject.

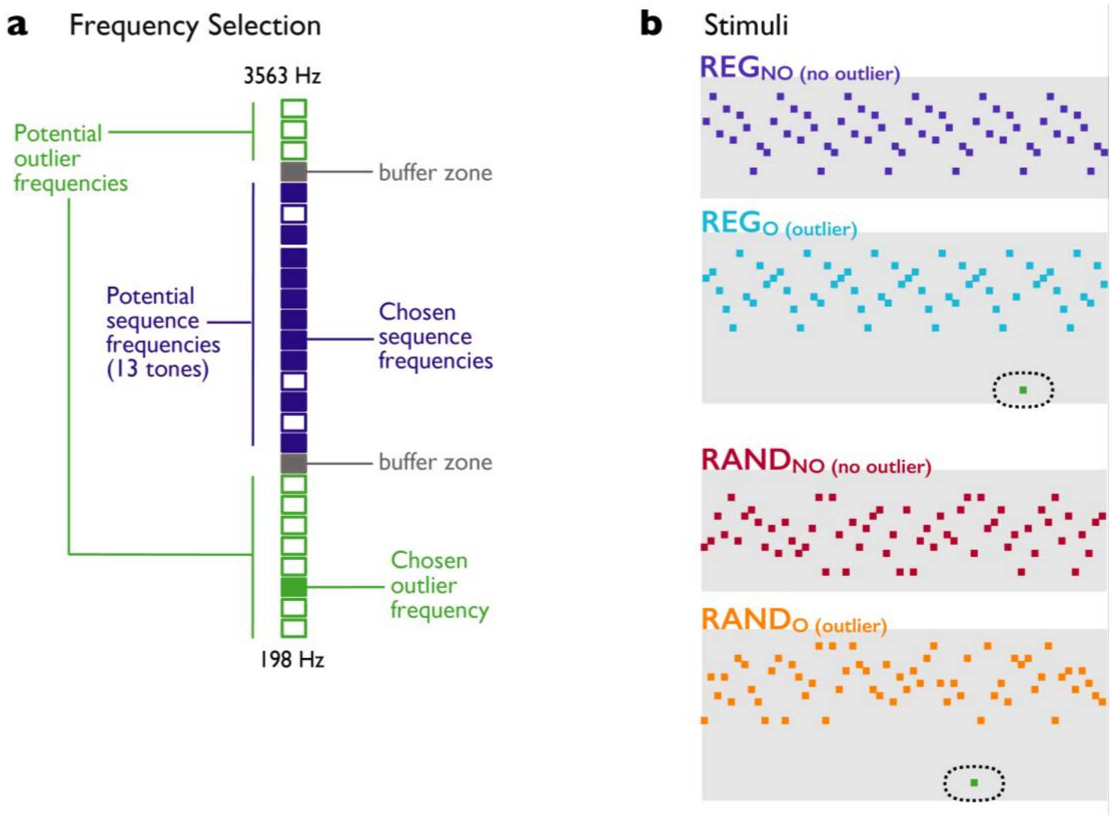

Figure 5.1 Stimulus generation
a: Procedure for selecting frequencies used for each stimulus. From the pool of 26; 13 adjacent values were chosen at random as candidate sequence frequencies (purple); 10 were selected for the sequence. Of the remaining tones; all except the frequencies closest to the sequence could potentially be outliers (orange); and from these a single value was chosen at random to be the outlier on that trial. **b:** Example set of stimuli for the four conditions; these were generated together from the same frequencies in order to match acoustic properties.

5.5 Experiment 1: EEG responses to frequency-outliers

5.5.1 Methods

5.5.1.1 Stimuli & procedure

The stimulus set comprised four sequence types: $\textbf{REG}_{\textbf{O}}$, $\textbf{RAND}_{\textbf{O}}$, $\textbf{REG}_{\textbf{NO}}$, and $\textbf{RAND}_{\textbf{NO}}$ as described above. These were presented to naïve, distracted listeners whilst their brain activity was recorded with EEG. Each trial was unique and sequences were generated anew for each subject. A total of 600 sequences were presented; 150 of each condition. The session was split into 6 blocks to provide breaks, each with 25 trials per condition presented in a random order. The inter-trial interval (ISI) was jittered between 1100 and 1500 ms. Stimuli were presented with the Psychophysics Toolbox extension in Matlab (Kleiner et al. 2007), using insert earphones with the volume set at a comfortable listening level. In order to capture automatic, stimulus-driven deviance detection processes, subjects watched a subtitled film of their choice during the experiment, with the audio muted. They were informed that there would be some sounds played during the session, and were presented with a single example of RAND_{NO} as a demonstration; but were instructed to ignore all sounds.

Following the session, subjects were asked the following questions about the sounds they heard:

1. During the EEG experiment, you heard some sounds. How distracting did you find them (1 = not at all, 5 = very distracting all the time)
2. Please describe the sounds briefly – what did you notice?
3. Did you hear any patterns in the sounds?
4. Did you hear any beeps that broke the pattern?

5.5.1.2 EEG recording and analysis

EEG was recorded using a 128-electrode Biosemi system (Biosemi Active Two AD-box ADC-17, Biosemi, Netherlands) at a sampling rate of 2048 Hz. Data were pre-processed and analysed using Fieldtrip (Oostenveld et al. 2010) toolbox for Matlab (2015a, MathWorks). Separate analysis pipelines were used to analyse the whole sequence response (time-locked to sequence onset) and the deviance response (time-locked to the onset of the outlier tone). All filtering was performed with a zero phase-shift Butterworth filter.

5.5.1.3 EEG preprocessing

After epoching (see below), epochs containing artefacts were removed on the basis of summary statistics (variance, range, maximum absolute value, z-score, maximum z-score, kurtosis) using Fieldtrip's visual artefact rejection tool. On average 5% of epochs were removed for each subject (range 0-10%). Artefacts related to eye movements, blinks and heartbeat were identified using independent component analysis (ICA). Any channels previously identified as noisy were not included in the ICA procedure.

To analyse the *sequence-evoked response*, data were high-pass filtered at 0.1Hz (third-order) and divided into 5000-ms epochs (with 1000 ms pre-stimulus-onset and 1000 ms post-offset). After artefact rejection, all data were resampled at 200 Hz with an anti-aliasing lowpass FIR filter, and baseline-corrected relative to the pre-onset interval. Missing bad channels were reconstructed as the average of their immediate neighbours. Subsequently the data were re-referenced to the mean of all channels, averaged over epochs of the same condition, baseline-corrected (200 ms preceding stimulus onset) and low-pass filtered at 30Hz (fifth-order) for plotting and analysis.

For quantifying the *outlier response*, data were high-pass filtered at 2Hz (third-order) and divided into 700-ms epochs, with 200 ms baseline and 500 ms following the onset of the outlier tone. The cutoff frequency of 2Hz was chosen to ensure that any differences in sustained activity between REG and RAND have been eliminated. Conditions without a violation (REG_{NO} and $RAND_{NO}$) were epoched relative to the average outlier timing; rounded down to the nearest tone onset, i.e. 2100ms. These were used as a baseline against which the responses the outlier tones were evaluated. Note that after high pass filtering there was no difference between the REG_{NO} and $RAND_{NO}$ sequences (see Figure 2a for illustration and below for statistical analysis). Subsequent analysis steps were identical to the one described for the whole sequence analysis (above).

For the *offset response* analysis, the sequence-evoked data were high-pass filtered at 2Hz, re-aligned into epochs (2800-3500 ms) and baseline-corrected based on the interval 2800-3000 ms. Subsequent analysis steps were identical to the one described for the whole sequence analysis (above).

5.5.1.4 Statistical analysis

To assess the response to the outlier tones ('main effect of deviance'), we collapsed across context and computed the difference between trials which contained and did not contain an outlier. Formally this is expressed as the contrast: $(REG_O – REG_{NO}) + (RAND_O – RAND_{NO})$. Fieldtrip's cluster-based permutation test, which takes spatial and temporal adjacency into account, was used to correct for multiple comparisons (Maris and Oostenveld 2007; Oostenveld et al. 2010). The significance threshold was chosen to control family-wise error-rate (FWER) at 5%. This defined three regions of interest (ROI) in time-channel space showing a deviance response. To determine how the deviance response is affected by regularity ('effect of regularity'), we calculated an orthogonal contrast of the deviance response magnitude by sequence type, $(REG_O – REG_{NO}) – (RAND_O – RAND_{NO})$, for each of the ROIs defined above. Statistical analysis was performed across channels using the same cluster-based permutation test described previously. The same statistical procedure was performed to verify that there was no residual difference in the responses to REG_{NO} and $RAND_{NO}$, ensuring that any effect on the deviance response reflects processing of the outlier tone rather than differential processing of the control condition.

The offset response was compared between REG and RAND (collapsed across outlier and no-outlier trials), across the whole scalp and offset epoch, using the same clustering approach as described above for the deviance response.

To characterize the overall sequence-evoked response to REG and RAND, the root mean square (RMS) of the evoked potential over all channels was calculated for each time point to give a time-series which reflects the instantaneous power of the evoked response. In the current data, as well as previous studies with similar stimuli, the sustained response is characterised by a large DC-like shift without zero-crossings, and with similar response dynamics in all channels; thus the RMS is a faithful representation of the dynamics in individual channels (Barascud et al. 2016; Southwell et al. 2017). The distribution of RMS across subjects (mean, standard error, confidence interval) was then estimated for each condition using bootstrap resampling across subjects (Efron and Tibshirani 1993) with 1000 iterations, for plotting of the group average response in Figure 5.2a. The significance of the difference in RMS between REG and RAND was assessed using the same cluster-based permutation statistics as for the deviance response, at each time sample, from sequence onset to 500ms following offset. T-tests (2-tail) were performed using t-statistics computed on clusters in time, and controlled for a family-wise error rate of 0.05 (Maris and Oostenveld 2007).

5.5.1.5 Source analysis

In the absence of individual structural scans, a head model derived from a template MNI brain was used (colin27; as included in the Fieldtrip toolbox) for which the volume conductance model was computed from MRI images using the Boundary Element Method (Fuchs et al. 2002). A triangulated cortical sheet with 5124 vertices was derived from this scan and used as the source model. Source inversion was performed on individual subjects and separately for each condition, using Minimum Norm Estimation (MNE; Dale et al. 2000), as implemented in Fieldtrip. This method simultaneously fits multiple dipole strengths over the whole brain, allowing any combination to be simultaneously active. The solution is constrained by selecting the combination of dipoles with the lowest total energy which adequately explain the sensor-level data. The distributed nature of the solution is particularly ideal for locating the differences between REG and RAND; as from previous work (Barascud et al. 2016) it is expected that the response to these sequences involves multiple sources.

Source activity was reconstructed over a time-window spanning 0-300ms relative to the onset of the outlier. Source data were then averaged within the time intervals 80-145 ms and 165-245 ms which correspond to the two ROI time windows in which significant effects were found in time/sensor space. Subsequently, *t*-statistic maps were computed, within each time window, for the *main effect of deviance*: (REG_O + $RAND_O$) > (REG_{NO} + $RAND_{NO}$), and the orthogonal *effect of regularity*: (REG_O – REG_{NO}) > ($RAND_O$ – $RAND_{NO}$). Data were interpolated onto an inflated cortical surface for visualisation (Figure 5.2e&f) and are presented using a threshold of $t = 2$. Because the contrasts are motivated by significant effects in the time domain, further statistical inference was not performed to avoid circularity (per Gross et al. 2013). Due to the limited precision afforded by the template-based source modelling used here, we discuss activation patterns in terms of general areas as opposed to specific MNI coordinates.

5.5.1.6 Participants

Data from 20 paid subjects are reported (age 19-32, mean 22.8 years. 9 female). None participated in the behavioural study (Experiment 2). One additional subject was excluded from analysis due to excessively noisy data.

5.5.2 Results

EEG responses were recorded to REG and RAND sequences (Figure 5.1) which occasionally contained a frequency outlier. Overall frequency occurrence statistics, taken over the sequence duration or over the entire experimental session, are identical between REG and RAND. The resulting effect is that the context offered by each sequence differs in predictability but not in frequency span. In order to capture automatic, stimulus-driven deviance detection processes, participants were kept naïve and distracted, watching a silent, subtitled movie of their choice.

5.5.2.1 Post-session reports

Following the EEG experiment, participants were questioned about the sounds presented. Nine out of twenty described hearing some kind of pattern in the sound, for instance ‘repetition’ and ‘alternating high and low sounds’, although these descriptions were usually quite vague, and when pressed to elaborate, none had noticed the distinction between REG and RAND trials. Thirteen subjects reported hearing occasional sounds which broke the pattern, or were otherwise distinctive; and when asked to elaborate, several specified that the pitch of the tones stood out as higher or lower than the rest. This shows that the outliers entered subjects’ awareness at least in some cases, although accurate description of the patterning of the sequences was much rarer. The mean rating given for how distracting the sound sequences were overall was 2.2 out of 5, range 1-4; indicating that subjects were moderately distracted by the sound sequences on average, but with considerable variability.

5.5.2.2 Sequence-evoked EEG responses

Sequence-evoked responses (Figure 5.2a) were analysed by pooling across conditions which contained or did not contain an outlier. The standard sequence of auditory onset responses is seen, followed by a rise to a sustained response that persists until stimulus offset. The topography of this response for both REG and RAND is similar to the N1 onset response, namely a fronto-central negativity (see inset topographies; Figure 5.2c). The response to REG was significantly greater than that to RAND, from 705 ms after onset until 440 ms after offset ($p < 0.001$, FWER-corrected). The response to REG diverged from RAND after just 4 tone-pips (200 ms) of the first repeated cycle, demonstrating that the brains of naïve distracted listeners are sensitive to sequence structure, discovering the regularity very rapidly (in fact, as early as expected from an ideal observer see Barascud et al, 2016). Overall this pattern of results entirely replicates previous work (Barascud et al. 2016; Southwell et al. 2017). However, the present stimuli are better controlled for effects of frequency-specific adaptation, by ensuring that REG and RAND have exactly the same frequency content; and by disallowing repetitions of the same frequency on two adjacent tone-pips.

The bulk of the analysis (below) is focused on understanding whether, in addition to these global effects of regularity on the responses to the sequence, responses to the outlier tones are also affected.

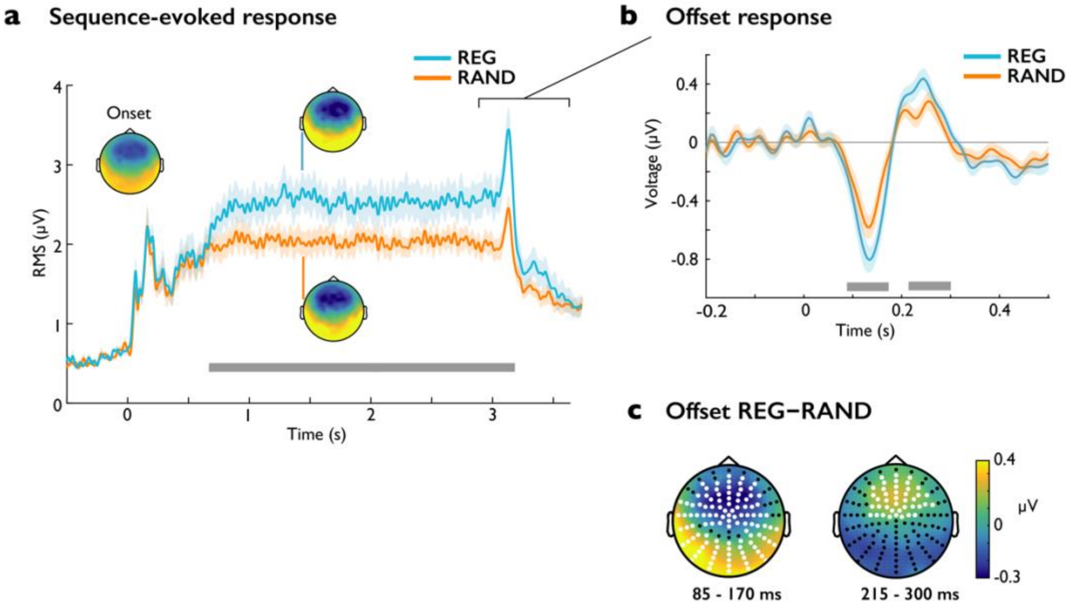

Figure 5.2 Sequence-evoked response
a: Sequence-evoked response. Shown in the main plot is the root-mean-square (RMS) of the signal over all channels, representing global field power; shading shows the standard error of the mean over subjects. Time period showing significant difference between REG and RAND conditions is indicated by a grey bar. Polarity-resolved topographies (across all channels) are shown for the onset response from 50-80 ms (inset; left) and the sustained response (700-3000 ms) to REG (inset; top) and RAND (inset; bottom). **b:** Offset response. Evoked response averaged over 58 central channels showing an effect of regularity on the offset response. Grey bar shows windows where there is a significant effect of regularity. **c:** Topography of the response during the two time-windows covering significant clusters for the contrast (REG–RAND); channels showing an effect of regularity on the offset response are highlighted in white.

5.5.2.3 Offset-evoked EEG responses

Interestingly, an effect of regularity is also present during the offset response, which is seen from about 50 ms after the cessation of the sequence (Figure 5.2b). The offset peak was compared between REG and RAND (pooling across trials which contained and did not contain a violation; high pass filtered to remove differences associated with the sustained response) using the same clustering approach as above. REG showed a significantly larger offset response than RAND (Figure 5.2c), from 85-175 ms ($p < 0.001$) in most channels (more negative in a fronto-central cluster of 58 channels, $p < 0.001$; and more positive in a temporal-occipital cluster of 50 channels, $p < 0.001$). There was also a significantly more positive response from 215-300 ms ($p = 0.008$) post-offset in a fronto-central cluster of 41 channels (Figure 5.2b; lower right). Statistical comparison was performed at each time-point and channel, but for illustrative purposes the time-domain response averaged over the 58 channels in the first negative cluster, is shown in Figure 5.2b.

5.5.2.4 Deviance-evoked EEG responses

For quantifying the deviance response (response to the outlier relative to the no-outlier conditions), data were high-pass filtered at 2Hz so as to remove the sustained response difference between REG and RAND sequences and focus on brain activity specifically evoked by the frequency outliers. A comparison between REG_{NO} and $RAND_{NO}$ confirmed no difference between these conditions after filtering.

The outlier-evoked responses (Figure 5.3a) were comprised of a series of peaks closely resembling the standard N1-P2-N2 sequence commonly observed at stimulus onset, or for changes within ongoing sounds (Martin and Boothroyd 2000). To quantify the effect of context on the response to the outlier, we first identified the channels and time intervals that show a response to the outlier (main effect of deviance ROI), we then investigated how this ROI is affected by context regularity (effect of regularity) by comparing outlier responses in REG vs RAND contexts.

To identify the main effect of deviance ROI; channels and time-intervals showing a response to the outlier, collapsed across REG or RAND context, were identified (see 'Methods'). This allowed separation of neural activity associated with the ongoing context of the sequence from those strictly evoked by the outlier tone. The resulting three ROIs, shown in Figure 5.3b, correspond to the peaks observed in the time domain (Figure 5.3a). **ROI1** comprised thirty-nine fronto-central channels which show a significant negativity between 80 and 145 ms (p = 0.001), corresponding most closely in time and topography to the N1. **ROI2,** a cluster of 33 channels at 165-245 ms ($p = 0.001$), had a similar topography but with a positive polarity, **ROI3**, from 290 to 320 ms ($p = 0.016$), had a smaller spatial extent (10 channels) and negative polarity (Figure 5.3c).

To quantify the effect of regularity on the outlier response, a comparison between deviance responses in REG relative to RAND was then calculated for each of the 3 ROIs identified above (see methods). In ROI1, a subset (21 channels) showed an effect of regularity on the outlier response (*p* = 0.005), which was 71% larger (calculated over mean activity within the significant channels), in REG sequences. In ROI2, responses were also larger (by 41%) in REG (*p* = 0.002) in a subset of 17 channels (Figure 5.3d). There was no effect of regularity in ROI3. Importantly, since the analysis above is performed on high pass filtered, and baselined, data, the effect of regularity on the deviance response occurs over and above the sustained response difference between the two sequence types.

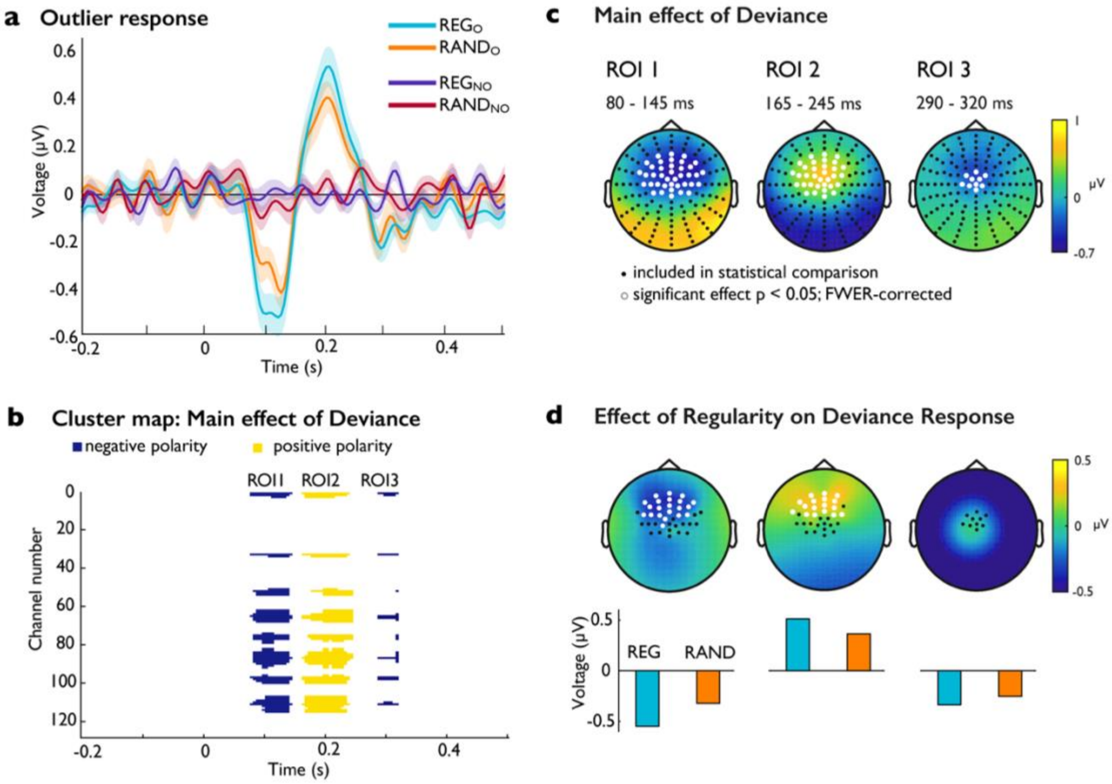

Figure 5.3 Deviance-evoked responses

a: Time-domain response averaged over the 39 central channels which showed a significant deviance response. Shading shows the standard error of the mean over subjects. The three deflections in the response correspond to the three clusters shown in (b). **b:** three time-channel clusters showing a main effect of deviance; i.e. (REG_O – REG_NO) + (RAND_O – RAND_NO). These channels and time-windows were used to derive ROIs 1-3. **c:** Topography of the three main-effect ROIs; averaged over the temporal extent of each cluster ROI. **d:** Topography of the effect of regularity: expressed by the contrast (REG_O – REG_NO) – (RAND_O – RAND_NO). Channels included in the statistical analysis are shown in black; these are the significant channels in (c). Channels showing an effect of deviance at any point during the cluster are highlighted in white. The average magnitude of the response to REG_O and $RAND_O$, within each ROI, is shown in the bar plots below.

Overall, the EEG results demonstrate that the brain rapidly detects the structure within REG and RAND sequences and is sensitive to the uncertainty induced by the sensory context, such that (frequency or offset) violations within a volatile (less predictable) RAND context are considered less surprising than identical events within a stable, predictable, background.

5.5.2.5 Source Analysis

Contrasts were also computed in source space (see methods), both for the **main effect of deviance,** and for the **effect of regularity.** The **main effect of deviance** in ROI1 was localised to bilateral temporal cortex (Figure 5.4a, top), maximal in right superior/middle frontal gyrus (S/MFG) with a peak *t*-statistic of 3.05. In ROI2, the main effect of deviance was associated with temporal lobe activation, but this time more prominently left-lateralised as well as situated more frontally around the temporal pole (TP), with a peak of $t = 3.55$ in the left middle temporal gyrus. Right-hemisphere activation is seen around the intraparietal sulcus (IPS) and the central sulcus (CS; Figure 5.4a, bottom).

For the **effect of regularity** (Figure 5.4b, top), in ROI1 we observed increased deviance response in REG at right TP and right orbital gyrus (OG), where the maximal *t*-statistic of 2.86 was observed. In ROI2, REG_O elicited a greater deviance response than $RAND_O$ in left temporal cortex, with a peak *t*-statistic of 3.05 in left middle temporal gyrus (MTG) and superior temporal sulcus (STS). Increased activity was also seen in right S/MFG.

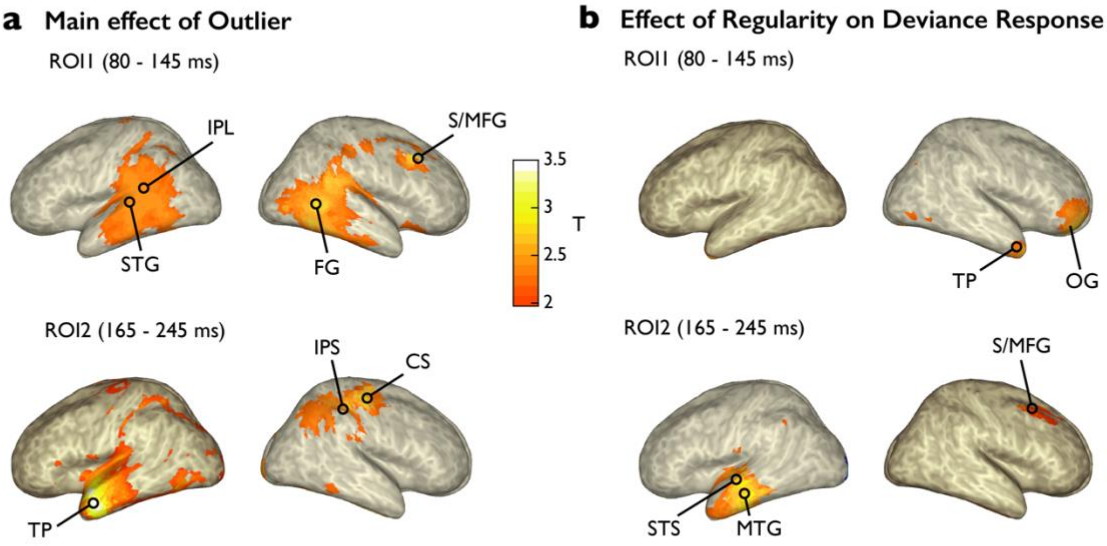

Figure 5.4 Source analysis

Source-level activity shown on a template cortical sheet. *t*-statistic maps thresholded at $t = 2$. All show average source activity taken over a time-window defined by ROI1 (80–145 ms) and ROI2 (165–245 ms) **a:** Main effect of deviance in ROI1 (top) and ROI2 (bottom). **b:** Effect of regularity on the deviance response in ROI1 (top) and ROI2 ms (bottom). Peaks of the *t*-statistic are indicated. Abbreviations: STG - superior temporal gyrus, IPL - inferior parietal lobule, FG - fusiform gyrus, S/MFG - superior/middle frontal gyrus, TP - temporal pole, IPS - intraparietal sulcus, CS - central sulcus, OG - orbital gyrus, STS - superior temporal sulcus, MTG - middle temporal gyrus.

5.6 Experiment 2: Behavioural detection of frequency-outliers

5.6.1 Methods

5.6.1.1 Stimuli & procedure

Subjects heard 96 trials each of $\mathrm{REG_{NO}}$, $\mathrm{RAND_{NO}}$, $\mathrm{REG_{O}}$ and $\mathrm{RAND_{O}}$ (in random order), and were instructed to respond by button press when they heard an outlier tone. Forty-eight additional control trials were also included, with the same number and timing of tone pips, but consisting of a single, repeating standard frequency (CTRL). Twenty-four of these contained an outlier tone at least 2 whole tones away from the standard ($\mathrm{CTRL_{O}}$); outlier and standard frequencies were chosen at random for each stimulus. Subjects were instructed to respond by button press as quickly as possible when an outlier tone was detected. Trials were presented in a random order, but the proportion of each condition across each block of 72 trials was kept the same. The testing session was preceded by a practice session of 28 trials; conditions were the same as the main experiment and in the same proportions.

5.6.1.2 Analysis

Dependent measures are d’ scores (Tanner and Swets 1954) and response times (RT; measured between the onset time of the outlier and the subject’s key press). Trials deviating from the condition-wise mean reaction time by more than 2 standard deviations were excluded; this resulted in exclusion of no more than 6% of trials for each condition. Sensitivity scores (d’) to outlier tones in each condition were calculated using the hit and false alarm rates. In cases where either rate was 0 or 1, a half trial was (respectively) added or subtracted to the numerator and denominator of the rate calculation; to avoid infinite d’ values.

5.6.1.3 Participants

10 paid participants took part (age 18-34, mean 24.4 years; 5 female). None participated in the EEG study (Experiment 1).

5.6.2 Results

We measured listeners' ability to detect frequency outliers in matched REG and RAND sequences (Figure 5.5). The mean reaction time to outlier tones in the control condition was 329 ms, giving an estimate of participants' basic response time and hardware/software latency. The average reaction times to REG and RAND were 347 ± 15 ms and 387 ± 25 ms; or 18 and 58 ms respectively when corrected for the basic reaction time. Reaction times to outliers in the control condition were significantly faster than both REG ($p = 0.012$) and RAND ($p = 0.0013$). When these reaction times are corrected for the basic response time, it is evident that detecting outliers takes only a few tens of milliseconds longer in complex sequences than it does in simple repetitive sequences. Paired-sample t-tests were carried out on the subject-wise averages of both RT and d' for REG versus RAND. Reaction times were significantly faster ($p = 0.01$) and sensitivity (d') significantly higher ($p < 0.001$) to outliers in REG, versus in RAND sequences.

To summarise, despite carefully matched properties of the regular and random stimuli used, we observe robustly greater behavioural sensitivity, as well as faster reaction times, to outlier tones which violate a regular sequence.

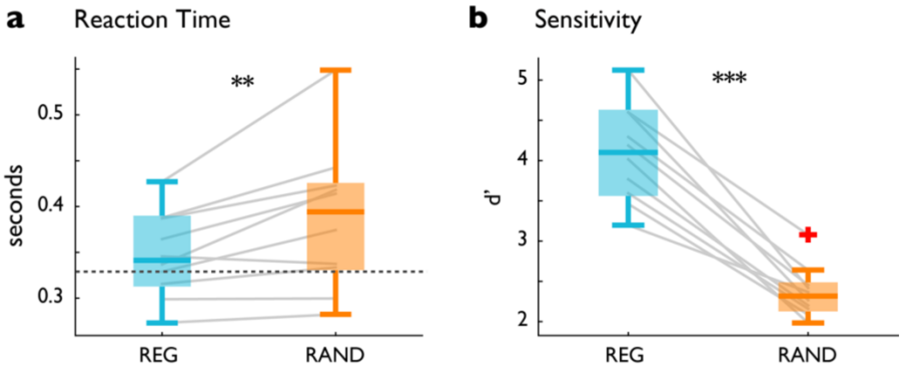

Figure 5.5 Behavioural results

Results from Experiment 2. **a:** Reaction times to outlier tones. **b:** Sensitivity (d') to outlier tones. **$p < .01$, ***$p < .001$. Individual subjects are shown with grey lines. Boxplots show distribution over subjects: median with lower and upper quartiles, whiskers show range. Extreme observations, falling outside of 1.5 times the interquartile range from the start of the whisker, are shown by a red cross.

5.7 Discussion

We investigated whether and how the predictability of successive events within rapid tone-pip sequences influences responses to outlier tones. Whilst it is commonly observed that regularity shapes responses to standards, even in complex sequences (Khouri and Nelken 2015; Heilbron and Chait 2017), effects on the response to the deviant outlier tone itself have been more elusive. For example, Yaron et al. (2012) report remarkable sensitivity to the temporal patterning of long sound sequences, but these effects are revealed via changes to the response to the standard, but not deviant sounds. Similarly, Costa-Faidella et al. (2011) showed robust effects of regularity on the standard, such that more repetition suppression is seen in a temporally regular than a jittered context - but the response to the deviant itself did not differ (see also Christianson et al. 2014).

5.7.1.1 Summary

Here, replicating our previous work (Barascud et al, 2016; Southwell et al. 2017) we observed substantial effects of context (REG vs. RAND) on the brain response to the sequence. Following the discovery of the regularity, REG elicited a higher sustained response. Importantly, we further demonstrate sizeable effects of sequence context specifically on the response to the deviant. Our results reveal two main findings: Firstly, robust effects of context were observed despite the fact that patterns were never repeated and had to be discovered anew on each trial. Though the outlier is set apart in frequency from the range defined by the sequence, and can in principle be detected based on this information alone, its detection was facilitated by sequence context. This was revealed in behaviour (Experiment 2) and in EEG responses from naïve distracted listeners (Experiment 1) where frequency outliers within regular sequences evoked a larger response (from 80ms after outlier onset) than matched outliers in random sequences. Secondly, the neural sources which underlie the effect of regularity, are, at least in part, distinct from those activated by the main effect of deviance (collapsed across REG vs RAND context). Whilst the latter was associated with the standard temporo-frontal network commonly implicated in frequency-based deviance detection, the effect of regularity was underpinned by sources in right temporal pole and orbitofrontal cortex.

The implications of these findings to our understanding of how the brain tracks and represents unfolding structure in rapid sensory signals are discussed, in turn below.

5.7.2 Automatic tracking of sensory sequence structure

Previous reports in the MMN (Paavilainen 2013, Bendixen et al, 2012), statistical learning (Koelsch et al. 2000) and music processing literature (Maess et al. 2001) have demonstrated increased responses to deviants within structured contexts, relative to random contexts. For example, in a study of musical expectation, Pearce et al. (2010) showed that low probability notes, compared to high probability notes, elicited a larger negative component at around 400 ms. Using non-musical, abstract tone sequences arranged in a random or ascending frequency pattern, Vaz Pato et al. (2002) demonstrated increased MMN responses to frequency deviants within the structured sequences. Koelsch et al. (2016) further showed increased negativity (from 130-220 ms post onset) to less probable items within sequences of tones with specifically controlled transition probabilities. Furl et al. (2011) trained participants to discriminate Markov sequences of pure tones from random ones and demonstrated a difference between low and high probability tones from 200 ms post-onset (during the P2 peak) originating in the right temporo-parietal junction.

However, a limiting factor in generalizing those results to listening in natural environments is the use of regularities established over an extended period. For example, a fixed pattern or transition probability matrix throughout the experiment; or even, for music, over a lifetime. As a consequence, these paradigms might be tapping long-term memory mechanisms; fundamentally different from those implicated in processing rapidly evolving and novel sensory sequences. Furthermore, brain activity was often recorded while participants were required to make decisions about the predictability of the pattern (Pearce et al. 2010; Furl et al. 2011) possibly implicating mechanisms related to active, overt tracking of sequence structure.

5.7.2.1 Sequence context is rapidly learnt and influences deviance responses

To probe rapid, automatic and pre-attentive processes associated with tracking evolving sensory statistics in the environment, we used rapid tone patterns (20Hz); beyond the rate which human listeners can actively track (Warren et al., 1991; Warren and Obusek, 1972). Unique sequences, whether REG or RAND, were used on each trial and (in Experiment 1) participants were kept naïve about the stimuli. We show that even when the regularity must be detected and represented afresh each trial, the response to a deviant is immediately modulated.

The deviant responses seen here - a standard succession of P1-N1-P2 deflections - are similar to those commonly observed in human literature (Briley and Krumbholz 2013; Herrmann et al. 2013) and which have previously been shown to be affected by both simple adaptation (repetition suppression) as well as more complex statistical context (relative probability of the deviant; Herrmann et al. 2013). Here we demonstrate a substantially larger response (71% increase in the first window) in REG relative to RAND sequences, confirming that these early deviant-evoked responses are also subject to automatic modulation by the degree of predictability in the ongoing sequence context.

In a separate experiment (Experiment 2), the effect of regularity was also revealed behaviourally: listeners are faster and substantially more accurate at detecting outlier tones within regularly repeating (REG), relative to random (RAND) tone-pip sequences, despite matched frequency content.

These findings are consistent with the notion that the brain continually tracks and maintains a detailed representation of the structure of the unfolding sensory input and that this representation shapes the processing of incoming information: deviants within high-precision sequences evoke higher prediction errors than identical events embedded in matched sequences of lower precision. A conceptually similar explanation may be framed in the context of perceptual binding: the tones in REG sequences are bound together by virtue of the underlying regularity model (Winkler et al, 2009; Andreou et al, 2011), such that deviants, not confirming to the rule, are perceptually represented as distinct 'objects' and therefore evoke a larger neural response. Alternatively, it could be the case that REG is easier to suppress and ignore, leaving more residual processing resources available to the detection of deviants. By 'subtracting away' the signals arising from the predictable sensory input, the deviant is relatively more prominent than in RAND. This hypothesis is supported by the behavioural findings in Chapter 3, where RAND was harder to ignore than REG. However, the present behavioural results cannot disambiguate these two accounts, and both may contribute.

Another alternative explanation for the observed findings might have been that regular patterns automatically attract attention (Zhao et al. 2013), and that this facilitates the detection of deviants in REG sequences. Chapter 3 directly investigated the question of whether attention is biased towards REG sequences (essentially identical to those used here), and found no attentional bias towards either REG or RAND. The fact that when interrogated, participants in the present study did not report noticing a distinction between REG and RAND trials also supports the conclusion that attention is not a likely explanation for the observed pattern of effects. Furthermore, the effects of attention on deviance detection are commonly associated with the presence of a P300 response (Chennu and Bekinschtein 2012; Molloy et al. 2015) reflecting the fact that the deviant was consciously perceived. The P300 was absent here. Instead our results point to an early and time-limited (between 80-250ms) effect of context on the deviant response.

5.7.2.2 Regularity increases the offset response

We also observed a remarkably strong effect of regularity on the offset response to the sequences. In the present paradigm, sequence offset is an instance of deviance, reflecting the violation of the expectation that a tone will be presented. This effect has been studied extensively in the context of the auditory omission (Chennu et al. 2016; Phillips et al. 2016) or offset (Andreou et al. 2015) paradigms, where an evoked response occurs to unexpected omissions of sounds, at a similar latency to the early responses to actual sounds, but only when the preceding sequence allowed a prediction to be formed about the omitted tone's properties. That both frequency and offset deviants are affected by regularity is consistent with the notion that the overall predictability of the pattern (the precision of the prediction the observer can make about an upcoming event) affects error responses regardless of the dimension in which the deviance occurs.

5.7.3 Source reconstruction

The main effect of deviance, computed by collapsing over sequence context and hence assumed to reflect the mismatch in frequency, was significant across a central subset of channels commonly associated with auditory responses (Figure 5.3c). In line with the standard network of bilateral auditory and right-hemisphere frontal sources often implicated in pre-attentive deviance detection (Opitz et al. 2002; Doeller et al. 2003; Garrido, Kilner, Stephan, et al. 2009; Halgren et al. 2010), source analysis suggested that activity within ROI1 (80-145 ms) originated in temporal cortex and right prefrontal cortex. later, in ROI2 (165-245 ms), the anterior portion of the left temporal cortex showed the strongest deviant-evoked response, with some additional activation in right intraparietal sulcus (IPS). The IPS is commonly implicated in auditory perceptual organisation (Cusack 2005) and specifically figure-ground segregation (Teki et al. 2016) and its involvement here may be linked to processes which stream the deviant tone away from the ongoing sequence.

The increased deviance response in REG sequences (effect of regularity) was associated with regions that are, at least in part, distinct from those involved in coding for the main effect of deviance. This was observed both in source space and in channel space, where the effect of regularity was only significantly present in a frontal subset of the channels identified as sensitive to the outlier.

In source space, the effect of regularity in ROI1 is underpinned by activity in the right temporal pole and right orbitofrontal cortex. This is in contrast to the main effect of deviance which is dominated by extensive activation of temporal areas. The right temporal pole and right orbitofrontal cortex have previously been implicated in sensitivity to context: the right anterior temporal cortex has been shown to be sensitive to the level of disorder in auditory and visual sequences, demonstrating higher activity the more ordered the sequence (Nastase et al. 2014). Orbitofrontal cortex has been proposed to be a source of top-down modulation on auditory cortex according to context (Frey et al. 2004) and is more generally implicated in integrating top-down priors with current information (Kepecs et al, 2008; Payzan-LeNestour et al. 2013; Wilson et al, 2014; Nogueira et al, 2017). The present results provide converging evidence for the role of these areas, outside of the standard deviance-detection network, in monitoring sequence structure.

Overall, source results replicate the ubiquitous network of bilateral auditory cortex and right pre-frontal sources as underpinning frequency-based deviance detection and additionally implicate the temporal pole as well as right orbitofrontal and pre-frontal cortex in nuancing these responses according to the preceding sequence context. This suggests that simple deviance responses are underpinned by activity in auditory cortex whereas more complex sequence structure related information is maintained outside of auditory cortex within frontal areas.

Source reconstruction based on EEG, particularly in the absence of individualised head-models, must be interpreted with caution. Future work, using more sensitive source-imaging, is required to understand and elaborate on these processes.

5.7.4 Implications for theories of predictive coding

All the deviant effects observed here were superimposed on an overall higher sustained response to REG relative to RAND patterns. A specific mechanistic account for the increased sustained response remains elusive, but previous work has demonstrated that the amplitude of the sustained response is related to the predictability or precision of the ongoing acoustic pattern (Barascud et al. 2016; Sohoglu and Chait 2016a; Auksztulewicz et al. 2017; Southwell et al. 2017), such that increased predictability is systematically associated with higher sustained responses. This effect, underpinned by increased activity in a network of temporal, frontal and hippocampal sources (Barascud et al. 2016; Auksztulewicz et al. 2017), may reflect a mechanism which tracks the context-dependent reliability of sensory streams.

Over and above this context effect, we demonstrated modulation of deviant specific responses. Though the present experiments do not provide evidence for a concrete link between the sustained response and the deviant response, they may be interpreted as reflecting two aspects of predictive coding. According to predictive coding theory, surprise is determined by two processes: *prediction error* evoked by a stimulus that differs from expectations, and also the *precision* associated with the input; i.e. the reliability attributed to the sensory stream (Kanai et al. 2015; Heilbron and Chait 2017). It is hypothesized that brain responses to predictable (highly precise) stimuli are up-weighted (e.g. through gain modulation) to focus perception on stable features of the environment (Feldman and Friston 2010). It is tempting to interpret the increased amplitude of the sustained response to regular sequences as a manifestation of precision-weighting (Barascud et al. 2016; Sohoglu and Chait 2016a; Auksztulewicz et al. 2017; Southwell et al. 2017), though it remains unclear whether the sustained effects seen here are indeed excitatory (as the gain modulation postulated by predictive coding; see further discussion in Southwell et al; 2017).

Importantly, the pattern of results we observe is not fully consistent with the standard predictive coding account of ‘prediction error’. Source analysis suggests the response to deviants in regular sequences was not merely enhanced relative to matched deviants in random sequences but rather arose in part via the involvement of distinct underlying sources. Therefore, an account in terms of differential precision weighing over the same prediction error units, as proposed by predictive coding (Feldman and Friston 2010; Kanai et al. 2015), may not fully account for the observed effects. Instead, the results point to a model where increasingly complex aspects of the same violating event are encoded in progressively higher stages of the processing hierarchy. In the deviance responses studied here this was revealed by predominantly auditory cortical sources coding for frequency deviance and frontal sources encoding more complex properties of pattern violation.

Chapter 6. Deviance Responses Beyond Outlier Tones

6.1 Summary

This chapter presents an EEG experiment and two behavioural experiments, measuring brain and behavioural responses to deviance from REG and RAND sequences. The previous chapter used tone pips at an unexpected frequency as the deviant event. The context provided by the predictability of the sequence was found to influence both behavioural and neural responses to deviant tones, which replaced the expected frequency of a tone with one at a higher or lower frequency than exists elsewhere in the sequence. This indicated that the precisely predictable context in REG led to up-weighting of the brain response to the unexpected tone, as compared to the equivalent outlier tone in RAND. This chapter uses silent intervals and noise bursts as deviant events instead, in order to further delineate the realm of influence that sequence regularity has on responses to transient events in the auditory scene. First, brief omissions were embedded in the sequence, to ask whether interruptions to the pattern *per* se are prioritised in a regular context, measured with EEG and behavioural studies. Second, I used a target detection task, with a noise burst superimposed on the sequence rather than replacing the tones, which was chosen to represent an acoustic event with maximally different physical properties to the REG and RAND tonal sequence. Whilst context regularity facilitated behavioural responses to omissions and noise bursts, there was no evidence that regularity affects the brain response to interruptions in the sequence.

6.2 Statement of Contribution

The work in this chapter was undertaken with a Masters student (Nicolas Abichacra; NA) under my supervision. Maria Chait (MC) and I designed the experiment. I collected the EEG data, apart from four subjects in Experiment 1 collected by NA. NA collected the behavioural data for Experiment 2. I wrote the code for stimulus presentation and analysed the data. MC and I interpreted the data. I wrote this chapter.

6.3 Introduction

The previous chapter asked whether the predictability of an ongoing sequence influences responses to outlier tones which replace the expected tone. Behavioural and EEG responses to the outlier tone, termed the **deviance response**, were both found to be enhanced by sequence regularity. In this chapter, the effect of sequence context on the deviance response is studied further, using different forms of sequence violation.

6.3.1 The brain responds to nothing

Another form of deviance from an established tone sequence is simply to omit a tone, violating the expectation that a particular tone, or simply any tone, would occur. Unexpected omission of an auditory stimulus has been shown to evoke activity closely resembling that of a veridical stimulus, originating from auditory cortex (Mustovic et al. 2003). For instance, Yabe et al. (1997) used isochronous sequences of identical tone pips presented at different rates, then occasionally omitted a tone. At sufficiently rapid ISI, below 150 ms, omissions were found to evoke a significantly larger response than to the ‘standard’ tone pips, peaking at a latency of around 120 ms; and this difference was described as the omission MMN. This effect was generalised to MEG by Raij et al. (1997), using similar but slower tone-pip trains with unpredictable omissions; the resulting omission-evoked field power peaked between 145 and 195 ms. More recent studies have shown omission-evoked responses as early as the P1 (Tervaniemi et al. 1994; Bendixen et al. 2009).

An omission within the context of the regular and random sequences in this thesis may be expected to produce such an omission response. This would violate the expectation of a tone occurrence, as well as the more specific expectation of a particular tone frequency which constitutes the expectation manipulation in the experiments in Chapter 5.

Omission responses have also been reported to occur in vision (Simson et al. 1977; Bullock et al. 1994) and somatosensation (Andersen and Lundqvist 2019). That there can be a response to an absence of incoming sensory information can be explained in terms of hierarchical predictive coding. By this account, evoked responses from a cortical region reflect prediction error, which is the discrepancy between actual afferent activity and the predicted afferent activity. Thus, the omission response magnitude, where the difference is between the prediction and zero afferent activity, is thought to mirror the strength of purely top-down sensory activity resulting from the expected sound (Wacongne et al. 2011).

6.3.1.1 Context effects on the omission response

The omission response is sensitive to the context in which the omission occurs. Several studies have shown that expected omissions evoke a smaller response than unexpected omissions (Todorovic et al. 2011; Wacongne et al. 2011; Chennu et al. 2016). In all these experiments, the expectation is manipulated block-wise whilst controlling for low-level stimulus features. This suggests the existence of a neural predictive model tracking what is more or less likely to occur, and weighting brain responses accordingly.

6.3.1.2 Omission response reflects prediction

In the omission response literature, as for the field more widely, evidence is thin on the ground for the deviance response as revealing a prospective prediction signal, as opposed to the possibility it results from a retrospective comparison. Bendixen et al. (2009) found omission-evoked responses to be modulated by predictability of the omitted tone within the first 50 milliseconds, specifically that the response resembled that to a real tone only if the tone was predictable from the preceding context. Chouiter et al. (2015) decoded, at greater than chance level, the expected direction of pitch change from the omission-evoked response, based on a classifier trained on the responses to real tones. It is also unclear whether the omission response, viewed through the lens of predictive coding, is directly measuring the activity carrying the prediction (e.g. as suggested by the modelling results of Chennu et al. 2016), or whether it is rather the prediction error resulting from its subtraction from zero afferent activity (as presented by Wacongne et al. 2012). However, in either case, predictive coding would hypothesise a differential weighting of omission responses in REG and RAND; either resulting from a stronger (top-down) prediction, or from a stronger (precision-weighted) prediction error arising from lower processing levels.

Alternatively, explanations of the existence of an evoked response to omission appeal to adaptation or rebound neural activity (May and Tiitinen 2001; 2010). In this model, neural subpopulations selective for regular temporal structure entrain to the stimulus presentation rate. At offset, these circuits continue to oscillate for a few cycles, producing a ‘rebound response’ which resembles the response to veridical stimuli. As discussed in more detail in Chapter 4 (§4.6.4), this model would also predict a larger omission or offset response in the case of REG than in RAND; because of the additional periodic structure present in REG at the cycle rate. However, such an adaptation model would not account for omissions evoking a larger response than the tone they replace: if anything, the effect would be opposite.

This co-existence of prediction-based and adaptation-based accounts of the omission response mirrors the debate on the cause of the MMN to oddball stimuli. However, it has been demonstrated that adaptation alone cannot explain the properties of the omission response in some cases. Wacongne et al. (2011) also manipulated the most likely identity of the tone, positioned at the end of quintuplets of tones, which was occasionally omitted: this could be either at the same frequency as preceding tones, or a particular ‘deviant’ frequency. Omission responses were larger in the latter case, whereas the adaptation model would predict equal omission response in both instances. Regardless of whether the process is retrospective or anticipatory, in the following EEG study, care is taken to equate the adaptive potential of the preceding sequences whether they are regular or randomly patterned. Therefore, the key novel comparison of the present EEG work – namely whether regularity modulates the omission-evoked response – should not be influenced by differential adaptation.

In EEG experiments in the preceding chapters, it was found that sequence offset is accompanied by a pronounced response peak, with a similar latency and topography to sequence onset and deviance responses. In fact, the offset response is modulated by sequence predictability in the same manner as the deviance response to outlier tones. Sequence offset is an extreme instance of tone omission; thus it may be that the offset response is itself a deviance response, showing precision-weighting according to the sequence context, of a prediction error resulting from the unexpected absence of sensory input. It is anticipated that an omission response, using the present REG and RAND sequences as context, will show a similar effect.

6.3.2 Outline

Experiments 1 and 2 below use the same stimuli as in Chapter 5 (see §5.4), however for REG_O and $RAND_O$, outlier tones are replaced by silent gaps. It is expected that such tone omissions will evoke responses similar to the onset, deviance and offset responses; namely a frontocentral negativity at around 100 ms, followed by a polarity reversal around 200 ms. Furthermore, this response is expected to be larger in the context of a REG sequence, which forms a more reliable (precise) context than RAND. Likewise, the behavioural sensitivity and reaction times to tone omissions is expected to be greater for omissions in REG than in RAND.

One further question which arises from the preceding experiments is that of the specificity of the mechanism prioritising deviant-evoked responses in REG. Is the contextual effect specific to similar tones which can conceivably be grouped with the ongoing sequence, or is it a more general effect, perhaps transferring to all auditory stimuli encountered in the same acoustic scene as a predictable sequence? Experiment 3 addresses this question, by measuring behavioural detection of a highly salient, nontonal target added to an uninterrupted REG or RAND sequence. Although target detection performance is expected to be at ceiling, it may be the case that REG still confers a reaction speed advantage, if the regularity of the context induces prioritisation of auditory inputs more generally.

6.4 Experiment 1: EEG Responses to Tone Omissions

6.4.1 Methods

6.4.1.1 Stimuli & procedure

The methods are as for Experiment 1 in Chapter 5 (see §5.5.1), except **REG$_\text{O}$** and **RAND$_\text{O}$** contain an omission of a single tone from the sequence, instead of a frequency outlier. Omissions occur with equivalent timing as in Chapter 5, i.e. locked to a tone onset between 1500 and 2750 ms, and consist of a 50-ms gap, which is the same duration as the missing tone (Figure 6.1).

Following the EEG recording, participants were asked the following questions about their perception of the stimuli:

1. During the EEG experiment, you heard some sounds. How distracting did you find them on a scale of 1 to 5 (1 = not at all, 5 = very distracting all the time)?
2. Please describe the sounds briefly – what did you notice?
3. Did you hear any patterns in the sounds?
4. Did you hear any breaks in the pattern?

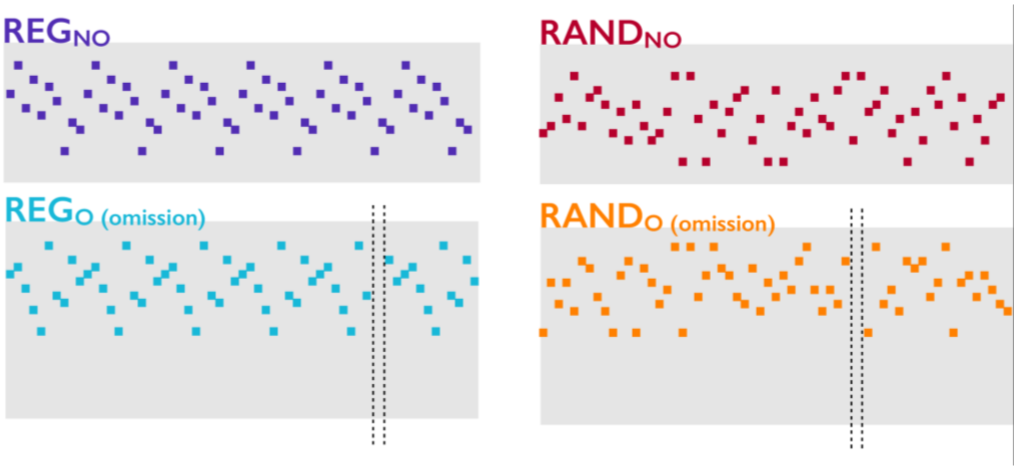

Figure 6.1 Stimuli
Schematic representation of sequences used in Experiments 1 and 2; with frequency represented vertically and time horizontally. In the latter half of the stimulus, an omission occurs in 50% of sequences (REG$_\text{O}$ and RAND$_\text{O}$). This omission lasts for 50ms, the duration of one tone, as indicated by the dotted lines.

6.4.1.2 Data analysis

Analysis of the EEG data was identical to that for Chapter 5, Experiment 1 (see §5.5.1), except no source analysis was performed due to the data being too noisy;

as determined by a lack of localisation of the auditory onset response to temporal cortex.

6.4.1.3 Participants

23 paid participants took part in the experiment. One subject was excluded from further analysis due to noisy EEG data. The final group comprised 22 subjects, age 19-27, mean 22.4; 8 males. None of the participants took part in Experiment 2 (§6.6 below).

6.4.2 Results

6.4.2.1 Post-session reports

Nine out of twenty-two described hearing patterns, although this recollection was usually vague, as was the case in the previous chapter. Only three subjects reported hearing the gaps. The mean rating given for how distracting the sound sequences were overall was 2.2 out of 5, ranging from 1 to 3; which is very similar to the ratings from Chapter 5 (§5.5.2).

6.4.2.2 Sequence-evoked response

The sequence-evoked response is shown in Figure 6.2. The REG sustained response diverged from RAND at 765 ms, but was only significantly higher from 1100 ms ($p < 0.001$ during the first cluster). The difference is significant at $p < 0.05$ throughout the sequence, apart from three brief gaps, until after offset.

6.4.2.3 Offset-evoked response

The offset response, analysed on high-pass filtered and baseline-corrected data to remove the preceding sustained response differences, is shown in Figure 6.2b-c. The offset response was larger for REG than for RAND. This effect was significant over most of the scalp, reflecting both the frontocentral-negative and corresponding positive temporal areas (see Figure 6.2c). There was a significant positive effect of regularity in 37 occipital and temporal channels from 105-150 ms ($p = 0.006$). There was a significant negative effect of regularity in 44 frontocentral channels from 100-155 ms ($p = 0.001$). This replicates the results from Chapter 5 (see Figure 5.3), but here was only significantly so during the offset N1, and for a briefer period than the 85-170 ms seen in the previous chapter, and here was also was not significant during the P2.

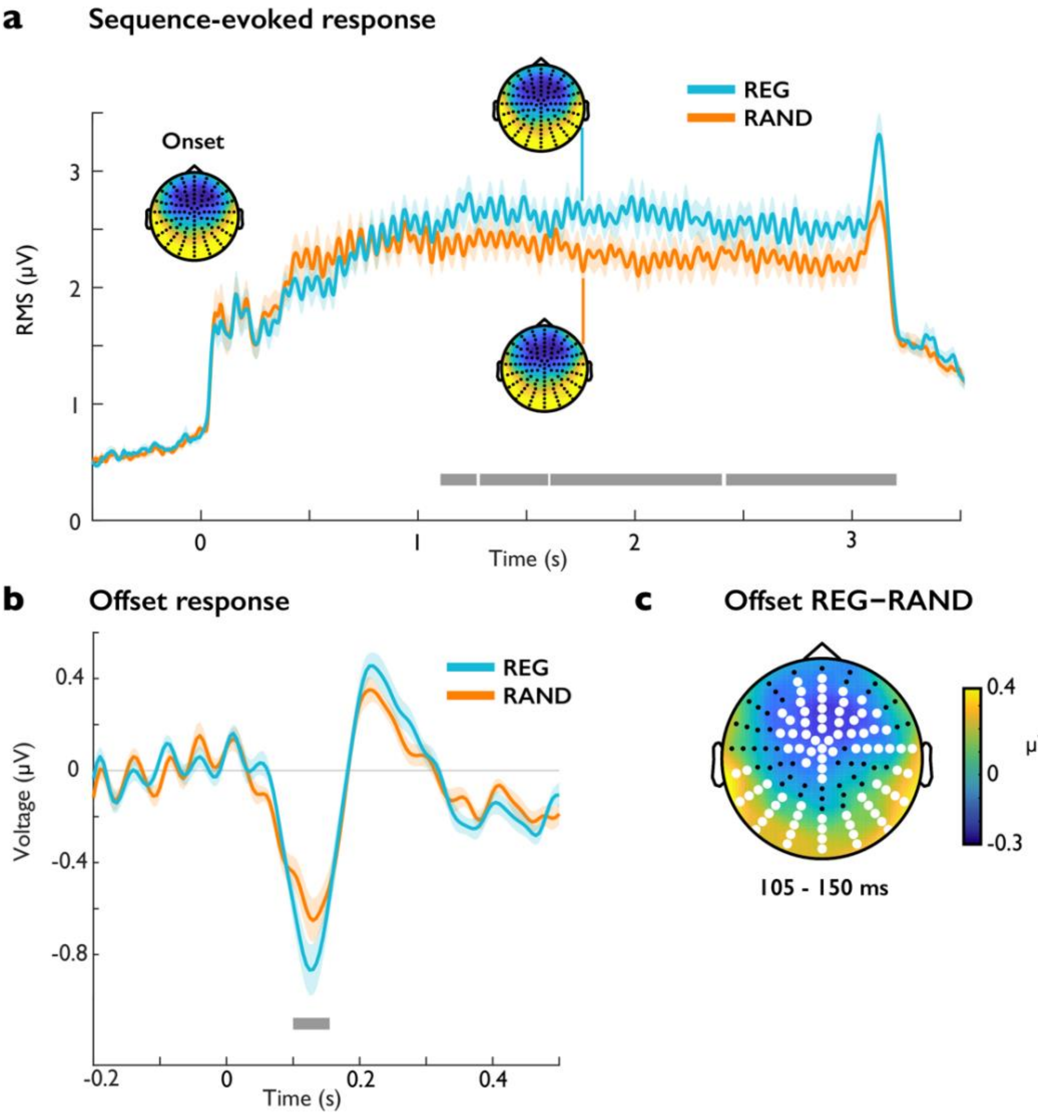

Figure 6.2 Sequence-evoked responses

a: Sequence-evoked response. Shown in the main plot is the root-mean-square (RMS) of the signal over all channels, representing global field power; shading shows the standard error of the mean over subjects. Time period showing significant difference between REG and RAND conditions is indicated by a grey bar. Polarity-resolved topographies (across all channels) are shown for the onset response from 50-80 ms (inset; left) and the sustained response (1100-3000 ms) to REG (inset; top) and RAND (inset; bottom). **b:** Offset response. Evoked response averaged over 44 central channels showing a negative effect of regularity on the offset response magnitude. Time period showing significant difference between REG and RAND offset is indicated by a grey bar **c:** Topography of the response during the time-window showing significant effects for the contrast (REG – RAND); channels showing an effect of regularity on the offset response are highlighted in white.

6.4.2.4 Omission-evoked response

The omission-evoked response is shown in Figure 6.3. There is an omission-evoked response which resembles the N1-P2 complex (Figure 6.3a), although this is smaller in magnitude than the outlier-evoked response in Chapter 5 (Figure 6.3a; plotted with the same y-axis scaling as in Figure 5.3 for comparison). The statistical analysis of the omission effect was carried out over all channels and time-points, with family-wise error rate controlled cluster-wise, revealing two significant clusters (Figure 6.3b). The time-averaged topographies within the time windows defined by the two clusters are shown in Figure 6.3c. Twenty-four fronto-central channels show a significant effect of omission from 90-100 ms (**ROI1**; $p = 0.008$) expressed as a more negative response. Fifteen central channels, mostly comprising a subset of those in ROI1, show a significant effect during the positive deflection of the omission response at 185-230 ms (**ROI2**; $p < 0.001$). The effect of regularity on the omission response was analysed using the contrast (REG_O – REG_{NO}) – ($RAND_O$ – $RAND_{NO}$). The statistical test was initially conducted within the two ROIs defined as a subset of time and of channels, but no significant clusters were found. Inspection of the topography of this contrast suggested that the difference between REG and RAND omission responses is maximal outside of the channel selection derived from the main-effect ROIs. Therefore the analysis was repeated over the whole scalp, within the same two time-ROIs, again showing no effect of regularity on the omission response (Figure 6.3d).

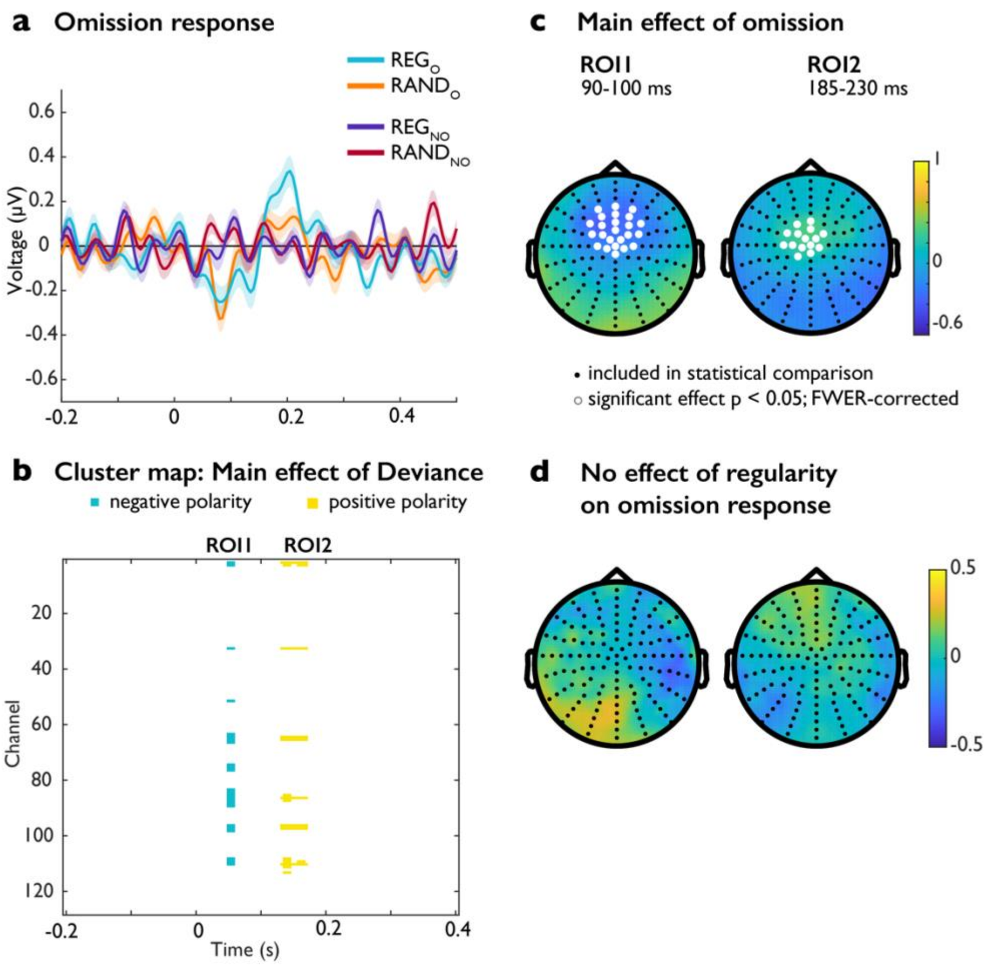

Figure 6.3 Omission-evoked responses

a: Time-domain response averaged over the 24 central channels which showed a significant omission response. Shading shows the standard error of the mean over subjects. The two deflections in the response correspond to the two clusters shown in (b). **b:** two time-channel clusters showing a main effect of deviance; i.e. (REG_O – REG_{NO}) + ($RAND_O$ – $RAND_{NO}$). **c:** Topography of the main-effect ROIs; averaged over the temporal extent of each cluster ROI. Channels showing an effect of deviance at any point during the cluster are highlighted in white. **d:** Topography showing contrast (REG_O – REG_{NO}) – ($RAND_O$ – $RAND_{NO}$). No channels show an effect of regularity on the omission response.

6.5 Comparison of Omission and Outlier-Evoked Responses

The EEG results from Chapter 5 (§5.5.2.4) were compared with those from Experiment 1 of the present chapter, in order to compare the omission-evoked and outlier-evoked deviance responses. The studies were performed with independent subject groups. However, the experimental design was kept virtually identical, aside from the nature of the sequence violation, in order to facilitate comparison.

6.5.1 Methods

The size of the deviance response in each experiment was summarised as follows. First, the mean of the evoked response for each subject was taken over a frontal subset of channels to obtain a timeseries for each subject. In order to minimise cancelling out electrodes with reversed polarity across the scalp, the *minimum* set of channels showing a deviance response in both experiments (main effect of deviance) was used; i.e. the intersection of the main effects of deviance across the two experiments. This was a frontocentral-negative selection of 24 channels for ROI1, and a frontocentral-positive selection of 15 channels for ROI2. These roughly correspond to a **deviance N1** and **deviance P2**. There was no significant omission response corresponding to ROI3 from the previous chapter.

The deviance response was calculated as the difference waveform REG_O minus REG; or $RAND_O$ minus RAND. To account for differences in latency & size of significant clusters, the peak magnitude of each subject's deviance response was calculated within two time windows chosen to correspond to the N1 and P2. The time range within which to search for peaks is taken as the union of the time-windows showing a main effect of deviance in the two experiments, i.e. 80-145 ms for ROI1 and 165-245 ms for ROI2.

The values were entered into a mixed-design ANOVA with a within-subjects factor of regularity and a between-subjects factor of experiment. A multivariate mixed-effects ANOVA was carried out with two separate dependent variables - N1 and P2 peak magnitude. Each dependent variable was modelled with a within-subjects factor (REG versus RAND). The experimental group formed the between-subjects factor (N = 20 for Experiment 1, N = 23 for Experiment 2). To clarify, whilst the multivariate ANOVA was run in one go, the magnitude of the peaks in the N1 and P2 time ranges were independently assessed and were not contrasted with one another.

From inspection of the omission-evoked response, it was expected that there would be an overall effect of group on the size of the deviance response, with smaller responses occurring in the omission experiment. As there was no effect of regularity on the omission response, but significant effects were found for the outlier response in both time-windows, it was expected that there would be an interaction between experimental group and regularity, indicating the regularity effect is significantly larger for the outlier tones than for omissions.

6.5.2 Results

Overall, there was a significant effect of regularity on the N1 peak magnitude ($F_{1,41}$ = 7.03, p = 0.011), where regularity overall served to increase the deviance response.

However, as reported in §6.4.2, there was not a significant effect of regularity on the outlier response just within the omission experiment. There was also a main effect of group ($F_{1,41}$ = 10.5, p = 0.002), with a larger deviance response to outlier tones than to omissions. However, there was no interaction between group and regularity ($F_{1,41}$ = .074, p = 0.79), so the size of the regularity effect was not significantly larger for outlier tones than for omissions.

For the P2, there was no overall effect of regularity ($F_{1,41}$ = 2.89, p = 0.097), however there was a significant effect of group ($F_{1,41}$ = 4.18, p = .045), with the outlier tones again evoking larger responses than the omissions. There was no interaction between group and regularity on the size of the deviance P2 ($F_{1,41}$ = 0.161, p = 0.69).

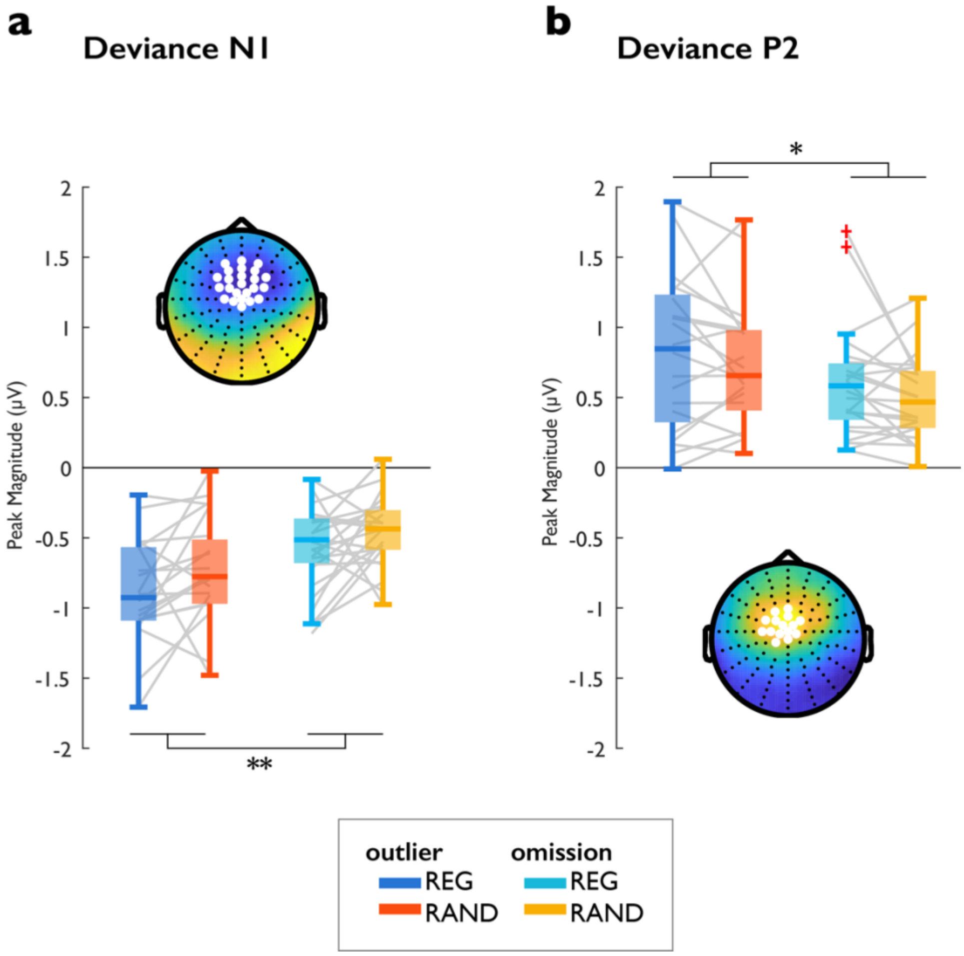

Figure 6.4 Comparison of outlier and omission experiments
Peak magnitudes of the deviance response, i.e. the difference between deviant and no-deviant waveforms averaged over the channels shown in the inset topographies. **a:** the peak magnitude of the deviance response in the N1 time window (80-145 ms). **b:** the peak magnitude of the deviance response in the P2 time window (165-245 ms). Figures a&b are shown with the same y-axis scale for comparison. Inset topographies show the channel selection over which each response was averaged before the peak magnitude was determined. Individual subjects are shown with grey lines. Boxplots show distribution over subjects: median with lower and upper quartiles, whiskers show range. Extreme observations, falling outside of 1.5 times the interquartile range from the start of the whisker, are shown by a red cross. * $p < .05$; ** p <.01

6.6 Experiment 2: Behavioural Sensitivity to Tone Omissions

6.6.1 Methods

6.6.1.1 Stimuli & procedure

The procedure was identical to Experiment 2 in Chapter 5 (§5.6.1), except 50% of sequence trials contained an omission in place of an outlier tone, as for Experiment

1 in the present chapter. Subjects were instructed to respond by button press when they heard an omission of a tone. Forty-eight additional control trials were also included, with the same number and timing of tone pips, but consisting of a single, repeating standard frequency (CTRL), 24 of which contained a tone omission at a time chosen from the same range as the omissions in REG_O and $RAND_O$.

6.6.1.2 Analysis

Analysis was identical to that for Experiment 2 in Chapter 5 (§5.6.1).

6.6.1.3 Participants

15 paid participants took part (age 20-32, mean 24.2; 8 female). None participated in the EEG study (§6.4). Three participants were excluded due to high false positive rates, indicative of misunderstanding the task and/or not concentrating.

6.6.2 Results

Experiment 3 measured the ability of subjects to detect tone omissions in matched REG and RAND sequences. The results, presented as mean reaction times and d' scores, are shown in Figure 6.5. The mean reaction time to omissions in the control condition was 378 ms, giving an estimate of participants' basic response time. Reaction times to omissions for the CTRL condition were significantly faster than in both REG (*p* < 0.001) and RAND (*p* < 0.001). The mean reaction times to omissions within REG and RAND were 463 ms and 545 ms, respectively. Reaction times were significantly faster (*p* = 0.0019), and sensitivity (d') significantly higher (*p* = 0.004) to outliers in REG, versus RAND sequences. As for outlier tones, the structure of the preceding sequence influences the detection of gaps within the sequence.

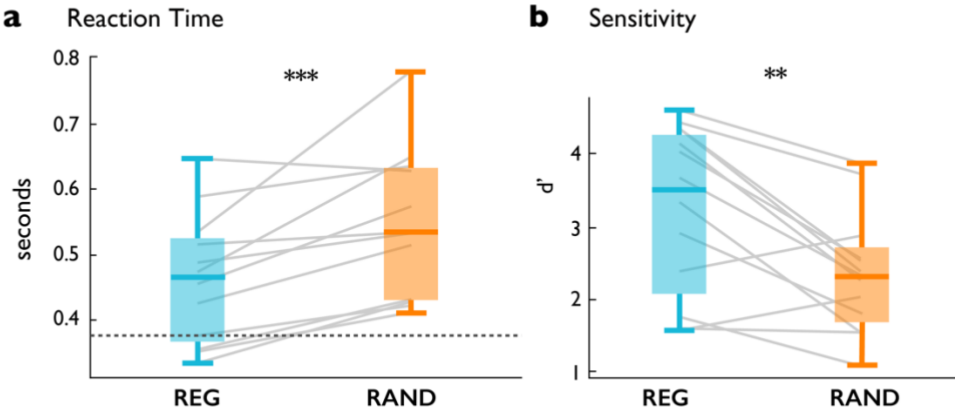

Figure 6.5 Omission detection

Results from Experiment 4. **a:** Reaction times to tone omissions. **b:** Sensitivity (d') to tone omissions. **p < .01, ***p < .001.

6.7 Experiment 3: Detection of Noise Bursts Within REG And RAND

The following experiment is similar to Experiment 2, but instead used a salient, non-tonal sound superimposed on the tone-pip sequence, rather than replacing the expected tone with an outlier tone. If regularity confers a general advantage to auditory target detection, then even detecting the appearance of a separate auditory object may be facilitated. The results of the following behavioural experiment using noise bursts instead of deviant tones, superimposed on REG and RAND signals, suggests the latter may be occurring, at least to some extent.

6.7.1 Methods

6.7.1.1 Stimuli

Stimuli consisted of three-second long tone-pip sequences with *alph* = 5, denoted **REG5** and **RAND5**. The sequences were generated from a frequency pool of 20 logarithmically-spaced values (222 to 2000 Hz; increasing by 12% at each step). Noise bursts were superimposed on 50% of sequences, with random timing, not constrained to coincide with tone-pip onsets, in the latter half of the stimulus; yielding a further two conditions indicated as **REG5n** and **RAND5n**. The noise bursts were band-passed (3-12 kHz) in a frequency range higher than the tone pip pool of 222-2000 Hz. Therefore, no predictions about tone frequencies were violated in terms of either a deviant frequency tone or an omission of an expected tone.

6.7.1.2 Procedure

The main experiment was preceded by a brief training block with six trials each of REG5, RAND5, REG5n and RAND5n. Subjects were asked to press the spacebar as soon as they heard the noise burst. The main block comprised 100 trials, with 25 trials of each condition. Subjects were given feedback on the screen in the form of a green tick or a red cross following each trial. To encourage rapid responses in the main task, the mean correct reaction time was taken from the training block, and feedback in the form of a yellow clock icon was given for any correct trials in the main task where the reaction time exceeded 200ms after this mean latency benchmark from the training.

6.7.1.3 Participants

10 paid participants took part (age 20-35, mean 25.8; 7 female).

6.7.2 Results

There was a small but statistically significant quickening of reaction times when the burst was superimposed on REG (Figure 6.6b). Hit rates for detection of the noise burst was very close to perfect performance for all participants (range 98-100%), and there was no significant difference in sensitivity to the noise burst as measured by d', between REG and RAND (Figure 6.6c). However, as the hit rates were so high overall, it is possible that the lack of effect is attributable to a ceiling effect. That there was some behavioural benefit of REG, or cost of RAND, even in such a simple, orthogonal task with performance at ceiling, suggests that the up-weighting of sensory signals in a regular auditory context may be non-specific, as opposed to only affecting stimuli sharing many perceptual characteristics.

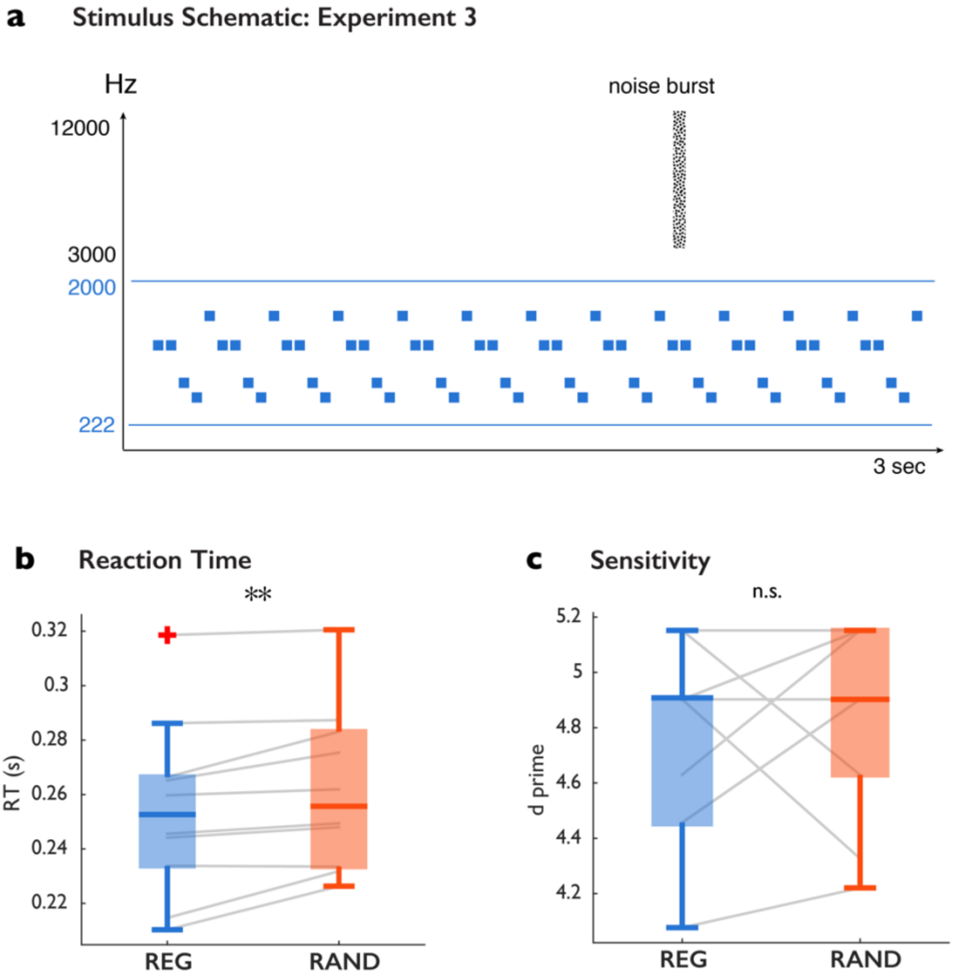

Figure 6.6 Noise-burst detection performance

a: Depiction of stimuli used in Experiment 3; shown is REGn, which is REG5 with a noise burst in the latter half of the stimulus. **b:** Noise burst detection as a function of sequence regularity. Reaction times were significantly faster to noise bursts in REG sequences. **c:** sensitivity (d') was not significantly different between REG and RAND contexts. ***p* < .01.

6.8 Discussion

6.8.1 Summary of results

The three experiments in this chapter add to the work in Chapter 5 in finessing the influence of sequence regularity on responses to transient, surprising events occurring during the sequence. From predictive coding accounts, and strengthened by the results in Chapter 5 which fit this account, it was hypothesised that responses to such 'deviant' stimuli would be heightened when they occur in a regular context. Experiment 1 used an omitted tone instead of a violating tone, and, whilst there was an EEG response evoked by the omission, this was not modulated by regularity. However, sequence offset was accompanied by a regularity-weighted offset response as for previous studies. Experiment 2 showed that regularity facilitates behavioural detection of omissions, using the same stimuli as the EEG experiment. Finally, Experiment 3 used a surprising noise burst superimposed on REG and RAND sequences, to help delineate the specificity of the influence sequence regularity can have on behavioural detection. Even though the acoustic properties of the noise burst distinguished it clearly from the tone-pip sequence, and it was superimposed on the sequence rather than interrupting it, there was a reaction-time advantage for REG versus RAND.

6.8.2 Omission response

Offset and omissions in RAND and REG sequences all produced a response in the N1-latency range, starting around 100 ms, as a symmetrical, fronto-central negativity in the EEG response. This timing and topography closely matches the properties of many omission responses found in MEG and EEG (Todorovic et al. 2011; Wacongne et al. 2011) and from intracranial recordings (Hughes et al. 2001). However the present omission effects were slightly earlier than the 150-200 ms found in other auditory EEG studies (Raij et al. 1997; Chennu et al. 2016), and considerably later than the 10-50 ms found by Bendixen et al. (2009). Variability in the properties of these omission responses may stem from the exact contrasts chosen to represent the omission response. Here, the main effect of omission is computed as epochs with an omission minus epochs without an omission; but which contain otherwise identical physical stimulus properties and statistical context. Therefore, the present omission response represents the prediction error, but any ongoing predictions are identical for the two conditions comprising the contrast. For several other studies, statistical analysis is focused on expected omissions minus unexpected omissions (Hughes et al. 2001; Todorovic et al. 2011; Wacongne et al. 2011; Chennu et al. 2016); thus the resulting evoked response also contains contributions from differing predictions, and therefore differing prediction errors; even though in this case the physical property of omission is identical for both terms. This may explain why the effects reported by such studies are sometimes later than the peak of the raw omission-evoked response: the measured prediction error may arise from a hierarchically later stage of processing, where expectations are formed that a violation (the omission) is to be expected.

There was no effect of regularity on the omission response in the present experiment. This is at odds with findings from similar experiments which manipulate the predictability of the tone-pip which is to be omitted, rather than the predictability of the omission itself. Ono et al. (2015) found omission responses to omissions of a note embedded in a predictable sequence to be larger than to omissions in randomly rearranged sequences, between 100 and 300 ms post-omission. This study differed from the present Experiment 1 in several respects, which may explain the difference in findings. Firstly, the predictable sequences were identical on every trial, allowing formation of an even stronger prediction of tone frequency than for REG. Secondly, their paradigm was active detection of omissions during the EEG recording; which is known to enhance contextual effects on the omission-evoked response (Chennu et al. 2016). Other experimental paradigms also showed modulation of the omission-evoked response according to the predictability of the tone which the omission replaces.

Bendixen et al. (2009) found that the omission response was significantly greater when the omitted tone identity was predictable from the preceding context, than when the omission occurred in place of an unknown tone. This effect was found as early as 10 ms post-omission, and when controlling for physical properties of the preceding stimulus. However a seemingly opposite effect was found by Wacongne et al. (2011), whereby omission of the final tone in a quintuplet is greater when the block-wise context leads to the expectation that the final tone will be different (‘XXXXY’ block) from the others, as opposed to blocks where the final tone was at the same frequency as the previous (‘XXXXX’ block). However, this is explicable as a larger evoked response is expected to the final tone in an XXXXY sequence than in an XXXXX sequence; as shown by countless studies using the oddball paradigm (see Chapter 1 for a more in-depth review).

6.8.3 Comparison to the outlier-evoked deviance response

The effect of regularity on the EEG omission response was not significant, but the results from chapter 5 showed regularity does significantly increase the response to outlier tones. However an ANOVA modelling both experiments together showed an effect of regularity on deviance responses overall, in the N1 range. Moreover, the interaction of group and regularity was not significant. These seemingly contradictory statistical results are however entirely possible within frequentist statistics. It is not necessarily the case that a difference in significance between two effects, when assessed separately, is accompanied by a significant difference between the two effects when they are directly compared (Gelman and Stern 2006). Nevertheless, it is difficult to interpret the results of this chapter holistically.

One possibility is simply that the data in the omission EEG experiment are of lower quality than in the previous experiments. Compared to the near-identical experiment in Chapter 5, the timing of the REG vs RAND sustained effect is much later in the EEG data for the present chapter, reaching significance only at 1100 ms versus 690 in the outlier tone experiment. In both cases, this arises before any omissions or outlier tones, therefore both experiments used identical stimuli up to this point. A lower signal-to-noise ratio in EEG data from this chapter could obscure the presence of small effects. Indeed, in the P2 window, the difference in the grand average between the omission response in REG does seem higher than its equivalent in RAND (see Figure 6.3a). Also, it has recently been shown in MEG that the omission-evoked response does not occur in all subjects, whereas other deviance responses are more ubiquitous in the population (Recasens and Uhlhaas 2017). Nevertheless, when attended, outlier tones were detected above chance level by all subjects.

6.8.4 Offset-evoked response

The offset-evoked response could be categorised as a special case of an omission-evoked response. Whilst it would be trivial for the listener to form an expectation that sounds ended after a duration of approximately three seconds, it is unlikely that the precise moment of offset would be expected, as temporal judgement accuracy begins to drop off by this duration (Drake and Botte 1993; Mates et al. 1994). Furthermore, even regularly-occurring deviants elicit an error response, as evidenced by the local-global paradigm, where predictable deviants or omissions nevertheless elicit an MMN (Wacongne et al. 2011; Chennu et al. 2016). This is explained in the context of predictive coding by appeal to predictions generated at different levels of the cortical hierarchy. In this case, the regularly-occurring deviant violates the expectation at a lower level, whilst being predicted by a higher level tracking non-adjacent dependencies in the sound. Therefore it is possible that sequence offset is surprising enough to elicit a deviance response, at least at a relatively low level of the auditory-perceptual processing hierarchy.

However, the mean offset response is much larger than the omission response: around -0.8 μV as compared to -0.3μV. Also unlike the response to omissions embedded in an ongoing sequence, the EEG offset-evoked response was modulated by sequence regularity in both Chapter 5 and the present Experiment 1. One potential explanation for these discrepancies is that sequence offset, effectively an omission lasting over one second, is a stronger violation than the brief 50-ms omission during the sequence. A stronger violation will lead to a larger evoked response, consequently with a greater signal-to-noise ratio with which to detect differences between conditions. This explanation would require that the evoked deviance response is in part computed retroactively, with the response compounding rapidly as the silence extends beyond 50 ms, in order to generate a larger response but still at the same 100-ms latency as the within-sequence omission condition. A consequence of the stimulus design is that omission-related responses may be contaminated by responses to the resuming of the sequence following the brief omission. Anything occurring much beyond 50 ms after the omission could in fact be due to the first tone post-omission. However the timing and topography of the omission response is similar to the outlier response, rather than 50 ms later. This would suggest that the deviance effect in the present EEG experiment does reflect the processing of the omission.

6.8.5 Regularity improves behavioural performance

Behaviourally, although EEG omission responses were not significantly affected, context regularity still conferred a relative advantage to detecting omissions, as shown by Experiment 2. Experiment 3 also showed that stimulus regularity decreases the response time to a reaction time task.

Improved target detection in REG could be explained as the result of the increased precision serving to up-weight prediction errors arising from the deviant sensory input. This up-weighting would increase the influence of the deviance on higher-order processes, including those related to forming the decision to respond. This account is supported by the fact the brain response to REG is consistently higher than to RAND, which lends more credibility to the hypothesis that behavioural detection is advantaged in REG contexts due instead to a prioritisation of neural processing which influences the processing of even unrelated deviants from the regularity.

Chapter 7. Changing Regularity

7.1 Summary

In previous chapters, the sustained response magnitude has been associated with sequence predictability. If the sustained response indeed reflects the inferred precision of ongoing structure in the auditory environment, then one can expect that response power will change in close step with any changes in the statistical properties *within* the sequence on a given trial. In this chapter, I use a statistical change-detection paradigm, in conjunction with passive EEG, to test this hypothesis. Each trial begins with a regular pattern, unique on each trial, which diverges into one of three possible fates: half of trials keep the same regular pattern, whilst the other half transition to either another regular pattern, or a matched random sequence. It is predicted that the transition will be associated with a drop in response power. On trials where the transition is to a new *regular* pattern, we expect the response to rise back up to the same level as the no-transition trials once the new pattern is learnt. When the transition is to a *random* pattern, the response power is expected to remain lower than on no-transition trials, reflecting the lower precision of the sequence. This effect was indeed seen in EEG, further corroborating that the sustained response is a correlate of the neuronal tracking of stimulus precision in the very recent stimulus past.

7.2 Introduction

Sustained responses to REG and RAND stimuli have previously been shown to track the level of predictability during stimulus sequences governed by a particular rule. However, so far, all the underlying rules have remained static throughout a single trial. If the sustained response truly signals ongoing predictability, then changes mid-sequence to the regularity rule should be signalled by a concomitant change in response power, reflecting the process of learning and then representing the new rule.

7.2.1 Statistical object boundaries in auditory perception

As discussed in §1.2, auditory objects are not necessarily delineated by clear frequency separation, but instead spectro-temporal statistical boundaries form an important cue in realistic auditory scenarios (reviewed by Schröger et al. 2013). For instance, the emergence of regularity is perceptually and neurally detectable as a separate auditory object from a random background (Teki et al. 2013). Even when the spectral content, and overall temporal envelope, remains constant, the temporal regularity of individual stream elements is perceptually salient and reflected in automatic brain responses (Sohoglu and Chait 2016a).

Barascud et al. (2016) measured MEG responses to transitions from RAND20 to REG10, and in the reverse direction. In both instances, the sustained response field power tracked the degree of predictability. For RAND-REG, the response increased, rising above that to matched RAND without a transition from a cycle plus six tones of the new regularity; a latency similar to the divergence seen in both MEG (Barascud et al. 2016) and this thesis for the divergence of REG10 and RAND10. When the data were high-pass filtered with a 2Hz cut-off, the effect was no longer present, suggesting the detection of the new regularity is associated with the steady-state response shift. In contrast, the REG-RAND transition was accompanied by an initial increase in power, at around 150 ms, which then showed a steep decline to a lower plateau; this peak remained in the high-pass filtered dataset. This initial peak was interpreted as a mismatch response, related to the detection of the violation of the existing pattern, which represents a prediction error signal that may trigger subsequent updating of the listener's internal model to learn the new RAND pattern. The lower predictability of the post-transition RAND was encoded in the decreased sustained response.

7.2.2 Expected and unexpected uncertainty

Two pattern-tracking processes are introduced by the transition manipulation. First, the pattern change represents a violation of an ongoing pattern (be it deterministically regular, or statistically defined). Second, the pattern change introduces a new pattern which can be extracted. The former process requires only the detection of a single violating event, whereas the second requires some integration over time, as demonstrated by ideal observer models (Pearce et al. 2010; Skerritt-Davis and Elhilali 2019). This explains why the REG-RAND transition evokes a sharp mismatch response within the first couple of hundred milliseconds, whereas RAND-REG evoked no such response, but rather was accompanied by a gradual rise becoming significant after 6 repeated tones.

A recent Bayesian modelling framework, which infers Bayesian surprisal over (auditory) sequences, replicates this dynamical asymmetry for REG-RAND and RAND-REG (Skerritt-Davis and Elhilali 2019). The difference in sustained response reflects the detection of expected uncertainty in the RAND signal as compared to the REG. The encoding of this expected uncertainty allows the brain to down-weight the influence of the noisy signal on learning and inference. However, the transition from REG to RAND represents a sudden increase in unexpected uncertainty. Unexpected uncertainty may explain the rapid mismatch response, and signalling of the newly-acquired expected uncertainty is encoded in the relatively sluggish drop in sustained response.

Note that in Barascud et al.’s (2016) MEG study, the *sustained* response drop from REG to RAND was only significant from around 250 ms; or 5 tones after the violation of regularity, which is a similar latency to the sustained response rise from RAND to REG. The authors interpreted this as indicating an inherent delay in the mechanism of downshifting precision weighting: an ideal observer model can detect the change from the first tone, yet this information did not appear to be used to inform the signalling of expected uncertainty. This would seem to indicate an *opposite* asymmetry in neural processing of changes in predictability, i.e. that within the expected-uncertainty-tracking mechanism, increases in expected uncertainty are delayed relative to an ideal observer, whereas transitions to a more regular scheme are represented at a latency comparable to the theoretical limit. This asymmetry would be in addition to the already-discussed asymmetry between initial rapid detection of violation (i.e. unexpected uncertainty, or prediction error), versus gradual detection of changes in predictability (i.e. expected uncertainty, or precision). Nevertheless, it is difficult to draw this conclusion when the MMN-like error response, which increases the RMS power, temporally overlaps with the change in expected uncertainty which would cause a decrease in RMS power.

Further work from this lab has investigated behavioural and neural responses to transitions between different statistically-defined auditory objects. The theoretical asymmetry of order-to-disorder versus disorder-to-order detection has been borne out in these results. Chait et al. (2005; 2007) found that changes in interaural correlation statistics of acoustic noise are detected faster for transitions from correlated to uncorrelated than from uncorrelated to correlated, and this was the case in both behavioural and neural responses. Furthermore, the neural sources underlying these responses were distinct for the two change directions. Similar results was found for transitions between random tone pip sequences and constant repetition sequences (Chait, Poeppel, de Cheveigné, et al. 2007) equivalent to REG1 in Chapter 4, and for alternating repetition sequences, i.e. REG2 (Chait et al. 2008).

Different neuro-modulatory mechanisms are proposed to account for the signalling of expected uncertainty and unexpected uncertainty (Yu and Dayan 2003). Recent evidence from pupillometry, which indirectly measures phasic noradrenaline release, provides compelling evidence that the change from REG to RAND is interpreted as unexpected uncertainty and is signalled by noradrenaline, whereas the reverse change has no discernible effect on noradrenergic activity (Zhao et al. 2018).

7.2.3 Aims & hypotheses

This experiment uses REG stimuli, which are assumed to evoke a high sustained response as in previous chapters, which contain a transition midway to a different sequence. Unlike previous work, a transition from REG to a different REG forms one of the conditions, alongside transitions from REG to RAND. This manipulation subtly changes the contextual information available to an (ideal) observer: even if a transition is detected, there remains uncertainty about what will follow: it could be a new regular pattern or it could be random. Only tracking the frequencies past the first (potential) cycle of the new rule will allow this inference. In contrast, one cynical interpretation of the REG-RAND transition effects seen by Barascud et al. (2016) is that the drop in response power was due to a disengagement of pattern-tracking mechanisms given it was known from context that any transition from REG was necessarily to a RAND.

The stimuli in this chapter are generated such that the alphabet of ten tones used on each trial remains constant, even if there is a transition. This allows us to test whether the REG-RAND transition-evoked mismatch response seen in MEG truly tracks the violation of the spectro-temporal sequence rule, or just the introduction of a new note, evoking a relatively disinhibited ‘oddball’ response. Indeed the same argument would apply to the existence of the deviant-evoked response in Chapter 5. Under attention, it is possible to distinguish between regular cycles of 10 tones with only a single adjacent pair swapped, even down to a duration of 40ms (Warren et al. 1991). Therefore it is plausible that the transition from REG to a different REG or RAND will be perceptually detectable.

7.2.3.1 Predicted results

The predicted pattern of results is as follows. (i) Transitions will evoke an MMN-like mismatch response, riding on the ongoing sustained response to REG. (ii) This response will be followed by a drop in sustained response power at around 250 ms. (iii) For transitions to another REG, the sustained response will increase to the same level as no-transition REG, and this will occur during the first repeated cycle. (iv) For transitions to RAND, the sustained response will remain at the lower level. (v) The offset peak will be larger for reg-REG than for reg-RAND.

7.3 Methods

7.3.1 Stimuli

The stimuli are based on REG and RAND with *alph* = 10. Stimuli were 4.5 seconds in duration, with a fifty percent chance of a transition occurring at 2 seconds, to a new pattern (Figure 7.1a). **reg-REG** transitioned to a new REG10 pattern and **reg-RAND** transitioned to a new RAND10 pattern. The remaining fifty percent of trials contained no transition (**REG**). As in Chapters 3 and 4, sequences consisted of contiguous 50-ms tone-pips with frequencies chosen from a pool of 20 values logarithmically-spaced between 222 and 2000 Hz. To ensure that the transition represented a change in the pattern properties, without also consisting a change in simple acoustic features, the alphabet of 10 frequencies chosen on a given trial remained the same post-transition. However the selection of 10 frequency values was refreshed on each new trial. A further constraint was placed on transition trials such that the first tone-pip post-transition violated the established regular pattern. This ensured that the effective violation of the existing regularity indeed occurred at exactly two seconds.

7.3.2 Ideal observer model

As for Chapters 3 and 4, the stimulus set was modelled using IDyOM to ascertain the information content of each tone (Figure 7.1b). Differences in information content between conditions were assessed for significance using the same cluster-wise corrected procedure as used for the IDyOM results in §3.4.1 and §4.4.3. In particular, it was expected that the transition would be associated with a sharp increase in information content, representing 'surprise' as a result of the violation of the regularity. Indeed, reg-REG and reg-RAND both diverge from REG at the first tone following the transition, and this difference remains until the end of the sequence. Secondly, the model appears to re-learn the regularity after the transition in reg-REG with the same promptness as the initial recognition of regularity at sequence onset (see Figure 3.2). The information content in reg-REG diverges from that of reg-RAND at 13 tones post-transition, i.e. at a cycle plus three tones.

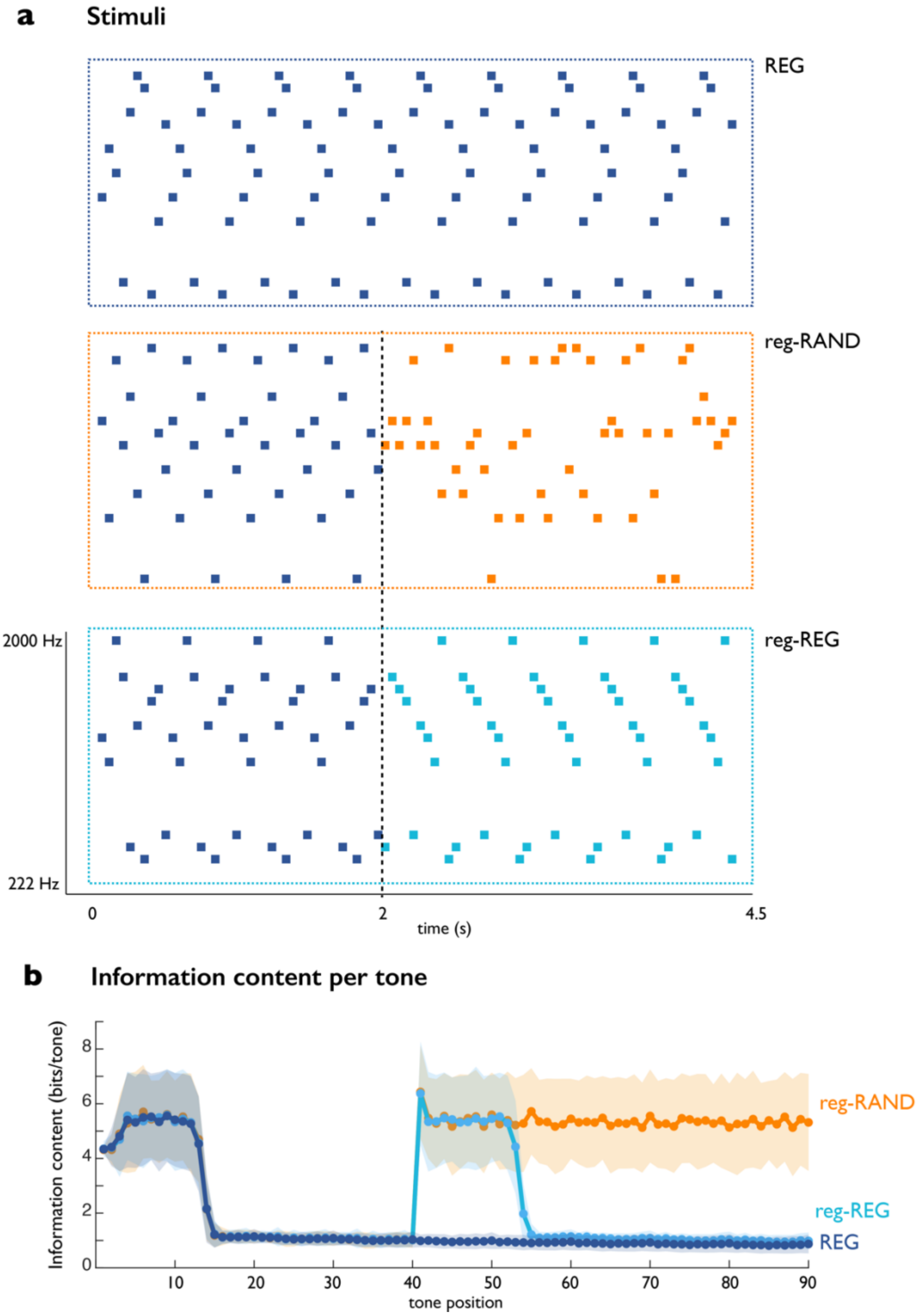

Figure 7.1 Stimuli
a: Stimulus schematics showing an example each of REG, reg-REG and reg-RAND. Squares represent individual tone-pips of duration 50 ms. The transition occurs at 2 seconds and is shown by a vertical dashed line. **b:** Information content per tone, estimated from a model of expectation computed over an entire experiment session. Mean information content computed over 128 trials of each condition. Shaded area shows standard deviation over trials.

7.3.3 Procedure

Each participant heard 128 trials each of reg-REG and reg-RAND, and 256 trials of REG, presented in a randomised order across eight blocks of 64 trials. Each block contained exactly 16 reg-REG, 16 reg-RAND and 32 REG trials. These proportions were chosen to render transitions and non-transitions equally likely, and that given a transition occurs, it could be to another regular pattern or a random one with equal probability. The inter-stimulus interval was randomised between 2 and 3 seconds. Stimuli were presented binaurally using the Psychophysics Toolbox (Kleiner et al. 2007) for Matlab, using insert earphones. Participants were instructed to ignore the sounds, and complete an incidental visual task, which was the same as that used in Chapter 3. The visual task was displayed on a separate computer using Cogent 2000 in MATLAB (www.vislab.ucl.ac.uk/cogent.php). The timing was not correlated with that of the auditory stimuli. For each trial, 3 colour photographs of landscapes were shown for 5 s each, and images faded gradually from one image to the next to minimise visual transients. Subjects were instructed to press a keyboard button if the first and third image within a trial were identical (10% of trials), and to withhold a response otherwise. The inter-trial interval was jittered between 2 and 5 s. Feedback (number of hits, misses and false alarms) for the visual task was provided at the end of each block.

Following the EEG session, subjects were asked the following questions, to ascertain what they noticed about the unattended sounds.

1. How distracting were the sounds? (1 = not at all, 5 = very distracting all the time)
2. What did you notice about the sounds?
3. Did you hear any patterns in the sounds?
4. Did you notice any sudden changes in the pattern?

7.3.4 Participants

Twenty four subjects took part in the experiment. Data from four subjects were rejected due to excessively noisy data. The final dataset of twenty subjects comprised 13 female, mean age 24.5, range 21-29 years.

7.3.5 Recording & data preprocessing

EEG signals were recorded using a 64-electrode Biosemi system at a sampling rate of 2048 Hz. Data were analysed using the Fieldtrip toolbox (www.fieldtriptoolbox.org/) (Oostenveld et al. 2010) for Matlab (2015a, MathWorks). Data were split into 8.5-s epochs, with 2 s before stimulus onset and after offset. An anti-aliasing 100 Hz low pass filter was applied before down-sampling at 200 Hz.

Detrending was applied to the data, to remove slow drifts in the EEG signal. A linear trend was fitted over each epoch; weighted such that only the signal before sound onset and 500ms after offset contributed to ascertaining the best fit (de Cheveigné and Arzounian 2018 see also Chapter 2). This minimised removing any slow trend in the data due to the auditory evoked response itself. This linear function was then subtracted from the data. Epochs were subsequently trimmed to contain only 500 ms before onset and after offset, then baseline-corrected by subtracting the mean signal in each channel over the 200 ms period before onset. Outlier channels and trials were then removed using Fieldtrip's visual artefact rejection tool, retaining on average 92.5% of trials for each subject. At most three channels per subject were rejected, out of a total of 64. ICA was used to detect and remove eye-movement artefacts (§2.3.4.4). DSS was employed to further remove noise by maximising repeatability across trials, retaining the top three components (de Cheveigné and Parra 2014; for more details see §2.3.5). Only the first 500 ms post-onset was used to determine the components maximising repeatability; this is because DSS will only detect repeatable activity which is exactly time-locked; and the temporal properties of responses on individual trials are likely to diverge later in the epoch. Finally, missing channels were reconstructed from the average signal in neighbouring channels, and baseline correction was re-applied.

7.3.5.1 Sequence-evoked response

For each subject, the evoked response was computed as the average over trials of each condition. For REG trials, as these were twice as prevalent as each of reg-REG and reg-RAND trials, half of the trials were randomly selected before averaging. The sequence-evoked response was summarised as the root-mean-square (RMS) power over all channels.

7.3.5.2 Transition-evoked response

An alternative analysis was carried out to measure if the first deviant tone after the transition triggered a transient deviance response such as those seen in Chapters 5 and 6. Full epochs were high-pass filtered at 2Hz before trimming to focus on the transition region, i.e. from 1.8 to 2.5 seconds. Baseline correction was reapplied, over the 200-ms period pre-transition. Then, transition (reg-REG, reg-RAND) and no-transition (REG) trials were separately averaged. Pooled transition conditions are denoted as reg-TRANS. As the regularity of the second portion of the sequence cannot be ascertained during the first ten tones post-transition, reg-REG and reg-RAND were not considered separately. The transition response was summarised for statistical analysis by averaging over a subset of ten channels, selected as follows.

The grand average response over all conditions in each channel was computed, then averaged over time for the time range 100-200 ms after to the transition. The ten channels with the lowest (most negative) value in this average were chosen to represent the entire transition evoked response, in the range 1.8-2.5 seconds, for each subject. This procedure ensured the resulting evoked response would be polarity-resolved, and thus the peaks comparable to deviance-evoked responses and the MMN.

7.3.5.3 Offset-evoked response

For the offset-evoked response, the same approach was used as in previous chapters. Full epochs were high-pass filtered at 2Hz, then cut to the duration 4.3-5 seconds (offset occurred at 4.5 seconds). Baseline correction was reapplied, over the 200-ms period pre-offset. The offset response was summarised for statistical analysis by averaging over a subset of ten channels, selected on the basis of all trials as for the transition-evoked response, except the time range for selecting the most negative channels was 50-150 ms post-offset. As for the sequence-evoked response, the number of trials in the REG condition was halved by selecting at random, before averaging over each of the three conditions.

7.3.6 Statistical analysis

The statistical approach was the same as in previous chapters. Statistical analysis compared the channel-averaged timeseries between conditions at each time point using a two-tail t-test. This was family-wise corrected for multiple comparisons at $\alpha = 0.05$ using a cluster-based permutation procedure (Maris and Oostenveld 2007).

7.4 Results

7.4.1 Post-session reports

The mean distraction rating was 2.5 out of 5. Fourteen of the 20 participants reported hearing cycling patterns in the sounds, of whom four noticed changes in the patterns.

7.4.2 Sequence-evoked response

Figure 7.2a shows the sustained response to all three conditions. All show the usual onset peak and rise to a sustained response. No conditions showed any difference prior to the transition, as should be expected. Around the transition, a small peak is seen in the transition conditions, but this was not significant. The sustained response power for both transition types drops below that of REG from around 500 ms post-transition.

Following the transition, both transition conditions show a drop in response power. For reg-REG, this occurred from 515 to 800 ms post-transition ($p = 0.0019$). Following the significant effect, reg-REG rises in response power toward that of no-transition REG. For reg-RAN, the response drops at 555 ms after the transition and remains lower than REG, though this is significant discontinuously (2.56-2.86 s, 3.31-3.47 s, 3.89-4.04 s, 4.43-4.67 s; all $p < 0.04$). Although reg-REG shows a higher response in the grand-average than reg-RAND from 2.80 s until offset, there is no *significant* difference between reg-REG and reg-RAND until almost offset (4.42-4.67 s; $p = 0.019$).

As predicted, there is a detectable decrease in the evoked response following transition to a new rule, as compared to a pattern that remains highly predictable. This is the case even though there are no changes in the constituent frequencies of the sequence.

However, the results only partly meet the hypotheses. Following the transition, it was expected that reg-REG and reg-RAND would diverge with much the same dynamics as REG10 and RAND10 in previous chapters. From the ideal observer model, reg-REG and reg-RAND diverge at a cycle plus three tones, corresponding to 2.65 s in the present stimulus. Whilst reg-REG diverges from reg-RAND in the mean at 2.8 s, which is a cycle plus six tones, the lack of a significant effect limits the conclusions that can be drawn from this. From onset, REG10 and RAND10 significantly diverged at around 0.9, 0.7 and 1.1 s in chapters 3, 5 and 6 respectively. In the present study, this corresponds to finding a significant difference from around 3 seconds from onset. There is no significant difference between reg-REG and reg-RAND until offset, so it is not possible to claim that the data truly show the processing of the re-emergence of regularity at the same latency as initial recognition of regularity. Nevertheless, each of the transition conditions separately shows the expected pattern of results in comparison to the no-transition control.

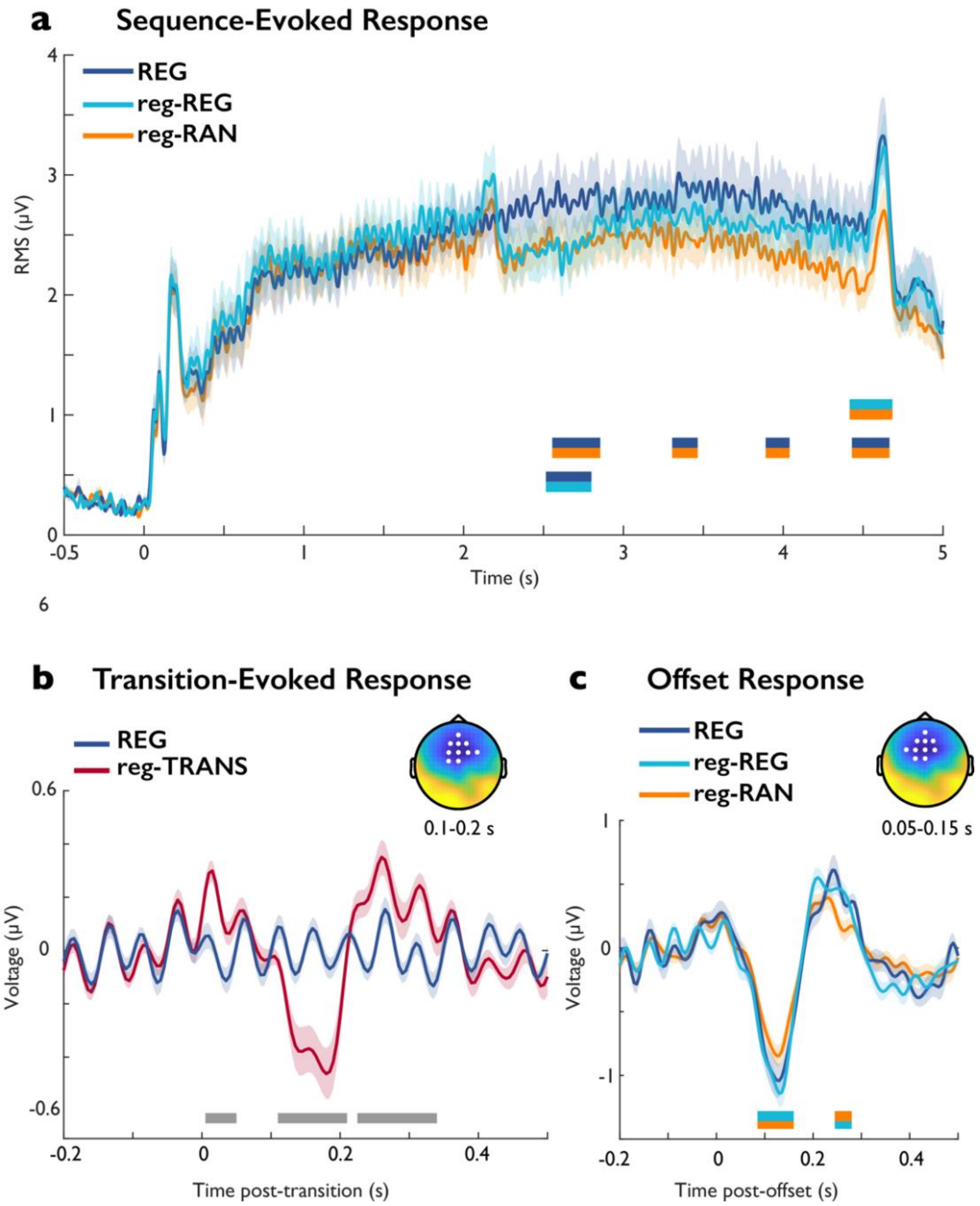

Figure 7.2 Evoked responses

a: Sequence-evoked response, RMS response power computed over all 64 channels. **b:** Transition-evoked response. Inset: topography showing response power at 0.1-0.2 s post-transition. **c:** Offset-evoked response. Inset: topography showing response power at 0.05-0.15 s post-transition. White circles denote the ten channels over which the average was computed. Shaded regions around time-series show the standard error of the mean over subjects. Grey or bi-colour bars under the evoked responses denote significant differences between pairs of conditions.

7.4.3 Transition-evoked response

The transition-evoked response is shown in Figure 7.2b. The topography shows the grand-average activity from 100-200 ms post-transition, alongside the ten channels over which the average was computed. This topography shows a frontocentral negativity, which is typical for auditory-evoked responses and the MMN. There is a transition evoked response resembling a P1-N1-P2 complex. There is a significant positive transition-evoked deflection from 5-50 ms, negativity from 110-210 ms, and positivity from 225 to 340 ms relative to the transition (all $p < 0.0064$).

7.4.4 Offset response

The offset-evoked response (Figure 7.2c) shows a typical negative peak around 100 ms and a positive deflection around 200 ms. The topography shows the grand-average activity from 50-150 ms post-transition, alongside the ten channels over which the average was computed. The offset response was larger for reg-REG than for reg-RAND, both in the negative deflection (110-160 ms; $p = 0.002$) and in the positive deflection (245-280 ms; $p = 0.025$).

7.5 Discussion

Brain responses were measured whilst the pattern underlying a regular sequence was changed, either to another regular sequence or to a random one, whilst maintaining the same spectral statistics. EEG responses showed an MMN-like peak immediately following the transition. This is thought to be a deviance response indicating the detection of a violation of the preceding regular rule. Unlike in Chapter 5, this violation consisted of a rearrangement of tones already existing in the sequence, rather than a new, extreme tone. This provides further evidence that the brain tracks ongoing pattern in sounds, signalling a prediction error when the pattern is broken. That the EEG evoked response differentiated between transition and no-transition trials from a mere 5 milliseconds is, however, surprising. Comparing this latency to other deviance responses in this thesis (and what we know about auditory evoked responses in general), it seems likely that the evoked response does not actually begin to differentiate between the conditions as early as this, potentially attributable to the use of zero-phase-shift filtering (see §2.3.4.1). The deviance response in Chapter 5, for instance, was significant only in the N1-like negative peak, from 80 ms.

After the transition to a random sequence, the sustained response power remained significantly lower than that of no-transition REG; this replicates previous findings in MEG (Barascud et al. 2016), but using a more pure manipulation of sequence statistics independent of frequency content. The novel transition type was from one regular pattern to another. This shows evidence of deviance detection, followed by a phase of learning the new pattern where the steady-state response is significantly below that of the unviolated sequence, before some tentative evidence that the response rises back to a similar magnitude as for the no-transition REG. The offset response is larger for REG and reg-REG than for RAND, replicating previous chapters. There is an additional hint that the ‘new’ regularity is represented as the same predictable ‘state’ as the unviolated REG control condition; because their respective offset responses do not differ; whereas the offset response for transitions to REG is larger than for transitions to RAND, even controlling for ongoing differences in the sequence-evoked response. In Chapter 4, the offset response also revealed sensitivity to patterning, in the absence of sustained response differences immediately preceding offset.

The data quality in this experiment may suffer from the increased influence that the EEG slow drift has on the longer trial lengths. Notably, more signal-to-noise reduction techniques were applied here than in previous chapters, yet there is still evidence of a higher noise level. Inspecting the sequence-evoked response, after offset, the power in the average response remains almost as high as the sustained response itself. In previous chapters, the sustained power decreased notably after offset. Reverting to using the same active visual task as in Chapter 3, rather than watching a subtitled movie, was motivated by aiming to reduce EEG noise from the stronger eye and face movements of the subjects. This appears not to have been effective.

According to the post-session reports, the number of subjects who reported noticing patterns, and changes in the patterns, was higher than in Chapters 5 and 6; indeed a majority described the cyclical nature of the patterns when pressed. This difference may be a result of the different decoy task used here: the visual task is potentially less engaging than watching a film, leaving more processing capacity available for consciously registering properties of the auditory stimuli.

The overall aim, and claim, of this experiment, was to demonstrate that the evoked brain response tracks predictability in sequences and updates this response as the predictability changes. Future work may go further, and manipulate expectations of the listener as to the possible statistical rules following the transition, through for example the use of a blocked design. This would allow delineation of the possible sources of top-down knowledge on representations of environmental predictability. Can, for example, explicit knowledge override automatic regularity-tracking responses? However, the current passive-listening EEG paradigm may not be sensitive enough to detect such effects within a reasonable experimental session. Instead, MEG might be preferable, as it does not suffer from the same slow-drift in the signal at a time-scale which interferes with the detection of the sustained response.

Chapter 8. General Discussion

8.1 Summary

8.1.1 Thesis in a nutshell

This thesis aimed to investigate the automatic tracking of predictable structure in the auditory scene. The model used for predictability was spectrotemporal patterning in continuous, rapid sequences of tone pips. The main measures of regularity-tracking processes were derived from EEG responses, but complementary behavioural paradigms were used throughout. Overall, the results provide further evidence that the brain automatically tracks the statistics of sequential events in sound, and that this also influences responses to change.

8.1.2 Experimental results

8.1.2.1 REG versus RAND shows increased brain response power

Throughout this thesis, the EEG results in every chapter provide further corroborating evidence that abstract regularity is encoded by a general, sustained increase in brain response, upon which the responses to individual tone pips ride. Cycle lengths of 3, 5, 10 and 15 tones all produce this effect: an increase in response power for REG as compared to matched RAND, distributed over the scalp similarly to auditory onset responses. For all comparisons except REG3 versus RAND3 in Chapter 4, the sustained response increase remained stable throughout the sequence until offset. Furthermore, the latency of the divergence between matched regular and random sequences reveals successively later pattern discovery for longer cycles. The mean latency of divergence between matched REG and RAND was around the middle of the first repeating cycle, which is close to the ideal observer-derived lower bound on discovering regularity of a cycle plus four tones. The significance of this effect was generally found to be later, suggesting that some subjects are slower to detect the regularity than others. The responses to pattern changes in Chapter 7 tell a similar tale. Changes to the regular pattern, whilst retaining the same frequency distribution, were signalled with an initial drop in sustained response power, which then increased only for transitions to a new regularity. This suggests that the brain constantly monitors the auditory environment for predictable structure, and encodes this in an automatic and obligatory manner in the magnitude of the sustained response.

8.1.2.2 Sequence complexity increased brain response power

Varying the alphabet size of random sequences affected the sustained response power in a systematic manner. Generally, the smaller the alphabet size, the larger the sustained response, as shown in Chapter 4 (with *alph* = 3, 5 and 20). This was less clear-cut in Chapter 3 (which used *alph* = 5, 10, 15 and 20). There was a main effect of alphabet size on mean sustained response power when all conditions were compared in an ANOVA, however the time-resolved analyses showed no such effect. From Chapter 4, it appears that larger alphabet size per se does not cause the reduced response, but that this effect is only seen for random sequences. There, REG3 and REG5 were not significantly different except for the initial portion of the sequence where the regularity begins earlier in REG3 than in REG5.

8.1.2.3 Increased responses not attributable to attention

Chapter 3 used several behavioural paradigms to indicate that the regularity effect, where regular sequences evoke a larger response than random ones, was not attributable to increased exogenous attentional capture by regular patterns. Instead, Experiment 2 in Chapter 3 showed that random distractor sequences were harder to ignore than regular ones.

8.1.2.4 Offset responses index predictability independently from sustained responses

The EEG responses to sequence offset were computed in the latter four chapters. All consistently showed an increased response at offset for regular compared to matched random sequences. One interpretation of this is that the sustained response tracks predictability continuously and this informs the response to unexpected events occurring in that context; sequence offset being one such unexpected event. However, there is evidence from Chapter 4 that the offset response is not accessing exactly the same regularity-tracking mechanisms as the sustained response. The offset response for REG3 was much greater than for RAND3, despite the preceding sustained responses showing no differences.

8.1.2.5 Deviance detection

Behavioural responses to deviant events occurring within a regular sequence are faster than equivalent events in random sequences, and these deviant events are associated with stronger brain responses in a passive listening condition. This is true for deviant tones, deviant noise bursts and sequence offset. This provides evidence that error-signalling is modulated by context, such that deviance from a more reliable rule is afforded more weight in perception and processing than the same physical event occurring against an uncertain backdrop. Source analysis tentatively suggests that the increased deviance response in regular sequences is at least in part due to the activation of additional sources, rather than just increased gain on responses within the same deviance-detecting sources. No such interaction between context and deviance response was found for EEG response to brief omissions within the sequence, despite the presence of a behavioural effect. Whether this reflects the elusiveness of the EEG omission response or a true qualitative difference in how deviance is signalled, is not conclusively revealed by the present results.

8.2 Implications

8.2.1 The brain automatically tracks rapid spectrotemporal structure

The sum total of the experimental results presented here converge on a view of the brain as possessing a powerful capacity for pattern recognition. The overall level of pattern is encoded, as well as individual elements of the pattern (20Hz response, deviance response, omission response, transition response) and furthermore, the sequence structure feeds back onto the responses to individual, transient events in the sequence (offset responses, deviance*regularity interaction in Chapter 5.)

This is consistent with a view of auditory scene analysis in terms of regularity extraction (§1.1.2, Winkler et al. 2009; Bendixen et al. 2010), and more generally with the brain as a predictive engine (Hohwy 2013; Denham and Winkler 2018). The REG/RAND paradigm allows the compression of a lot of spectrotemporal structure into just a few seconds, whilst being deterministic enough that changes and violating events are well-defined and well-controlled.

8.2.2 Evidence for Predictive Coding

One of the more specific motivations of this thesis was to test predictions derived from generalised predictive coding about the effect of uncertainty on brain responses. This is rendered problematic by the fact that subtly different interpretations of how the experimental manipulations map onto cognitive variables can yield sometimes opposing effects.

For instance, the main manipulation used throughout, that of deterministic sequential regularity, implies both increased precision and decreased prediction error. Precision is thought to be implemented by increased gain on the very same units that report the decreased prediction error; therefore an explanation exists for both increases and decreases in activity. Most previous auditory studies have shown repetition (or expectation) suppression, when interpreted through the predictive coding lens, this is portrayed as due to suppression by top down predictions (Garrido, Kilner, Kiebel, et al. 2009; Bendixen, SanMiguel, et al. 2012; Auksztulewicz and Friston 2016). The present results, however, show an increase in brain responses concomitant with predictability. This effect has two possible explanations consistent with predictive coding: increased gain, and increased top-down (possibly inhibitory) prediction activity (such as shown in Auksztulewicz et al. 2017).

The EEG experiment in Chapter 5 can potentially disambiguate these two possibilities. It was hypothesised that context regularity would lead to increased precision-weighting on bottom-up prediction errors arising from a deviant frequency. The results were consistent with regularity-induced gain, as equivalent deviants elicited a greater response when presented in REG, even when subtracting the ongoing sustained response differences. It is unclear how to account for this in a model where the regularity effect on the deviant is simply increased prediction. However, this does not conclusively demonstrate that the sustained response itself is due to precision weighting, as it remains possible that the regularity-induced deviant and sustained response increases are due to two different mechanisms. The picture is complicated by the source results, which show the regularity effect on the deviance response to be explained best by the involvement of distinct sources from those signalling the prediction error (main effect of deviance).

Neither does an appeal to predictive coding answer the question of why the sustained response increase is seen here, but not in other auditory experiments which manipulate sequence predictability. Why would precision-weighting or prediction dominate the EEG evoked response here, but expectation suppression be the dominant effect measured using the same technique but for different auditory stimuli? It is possible that both repetition suppression and repetition enhancement are present in both flavours of auditory paradigm, but the complexity of the current patterns favours a net response enhancement due to the involvement of a wider network of areas generating the predictions, for instance.

8.2.2.1 An attempt at synthesis

The following presents an attempt to unify the EEG results from the previous five chapters in terms of hierarchical predictive coding network. Although it is beyond the scope of this thesis to argue for this explanation to the exclusion of all others, it simply intended to provide motivation for further work to test the hypotheses arising from it.

At sequence onset, within the first 200 ms or so, the dominant effect on evoked responses is repetition suppression due to adaptation in primary auditory cortex (and lower-level areas, though these are likely to contribute less to the scalp EEG), causing decreased responses to subsequent tones occurring at the same frequency in quick succession. Repetition at a given frequency is most likely (and most frequent) in REG1, then RAND3, and as *alph* increases, the expected interval between identical frequencies becomes longer. Therefore, stronger frequency-specific adaptation would be expected the smaller the alphabet size (Ulanovsky et al. 2004; Yarden and Nelken 2017). This explains the reduced response at sequence onset to REG1 and RAND3 relative to RAND20, as seen in Chapter 4. This adaptation leads to reduced efferent activity reaching non-primary auditory areas, which influences recurrent message-passing between auditory cortex and higher-order areas. The higher-order areas attempt to explain away the reduced ascending prediction error, which favours a generative model engendering the expectation of recurrence of the same tone-pip frequency in REG1 and to a lesser extent RAND3, as compared to larger RAND alphabet sizes where repetitions occur less frequently and with a greater temporal interval. Over time, expectations of precision are also optimised; leading to an expectation of high precision associated with the ascending prediction errors the smaller the value of *alph*. This expected precision is implemented as an increased gain on the units reporting prediction error from tonotopic regions, the smaller the alphabet size.

A similar effect, operating over a slower timescale, would account for the REG versus RAND effect. Regularly-cycling temporal patterns are detected in prefrontal areas which are sensitive to more abstract, sequential relationships in the sequence of incoming prediction errors (Zatorre et al. 1994; Doeller et al. 2003; Overath et al. 2007). These areas send descending predictions, which have a net inhibitory effect, to the auditory cortex. This allows further reduction of ascending prediction errors from auditory cortex; in other words, expectation suppression. An expectation of high precision would be inferred from the consistent lack of prediction errors arising in REG sequences. This expectation of high precision would be fed back to lower areas as a modulatory influence, serving to increase the gain on the prediction-error reporting neurons in primary auditory areas. Sequence prediction-related activity, unique to REG, and increased precision weighting both lead to increased evoked response power to REG when measured over the whole scalp. An explanation of the REG versus RAND effect consistent with predictive coding would require that enhancement from precision-weighting, and from activation of additional neural populations participating in prediction signalling, outweighs suppression from reduced prediction errors as a result of the successful predictions in REG.

The increased gain on ascending prediction errors from tonotopic regions would explain how deviants and indeed offsets in REG elicit higher precision-weighted prediction errors. Adaptation caused by neuronal refractoriness is ongoing throughout the sustained response, but is mostly offset by other processes, except in the case of REG1, and later REG3, where this adaptation serves to reduce the effect of increased precision and the formation of a model of stimulus regularity outside of auditory cortex. At offset, the evoked response is a precision-weighted prediction error, where the error in question refers to the violated expectation that a sound will occur. The precision-weighting is largest for REG, followed by RAND in increasing order of *alph*.

Changes in the sequence of frequencies leads to greater ascending prediction errors; as seen as the main effects of deviance- and omission- evoked responses in Chapters 5 and 6, and the transition-evoked response in Chapter 7.

8.2.3 Entrainment: an alternative explanation?

Entrainment to structure in auditory signals, particularly at the sub-10Hz rate of syllables in speech, has been intimately linked to facilitated perception (Ding et al. 2017), memory (Wang et al. 2018), attention (Calderone et al. 2014), and auditory scene analysis (Riecke et al. 2015). This is the same range of rates at which cycles repeat in the present work, from 1.3 Hz in REG15, to 6.7 Hz in REG3. The strength of delta-phase consistency (1-4Hz) was recently shown to be correlated with spectral regularity in direct recordings from multiple cortical regions (Gifford et al. 2018). Oscillatory entrainment to REG cycles could explain the behavioural detection of regularity (Barascud et al. 2016) and potentially the facilitated detection of deviants in REG.

However this explanation is not entirely able to account for the EEG effects. For instance, entrainment models would not predict the increased deviant-evoked responses in REG as compared to RAND, and specifically do not account for the offset latency being similar for different cycle lengths (see Chapter 4 for further discussion). Also, entrainment during the sustained response was measured in Chapter 4 as indexed by the cycle-rate induced response. There was indeed an effect of regularity, but the effect size was substantially smaller than the sustained response RMS increase. The sustained response and oscillatory entrainment to similar stimuli has recently been shown to be dissociable under attentional manipulation (Herrmann and Johnsrude 2018). Barczak (2018) measured the local field potential (LFP) in monkey A1, in response to a continuous stream of alternating REG and RAND segments. They found increased delta-phase entrainment during pattern segments at multiple stages of the auditory pathway, with a top-down directionality of influence suggesting the entrainment represented pattern-recognition rather than a passive process. Although they set out to replicate the sustained response effect, they found the opposite: a reduction in the local field potential during regular segments.

8.3 Methodological Considerations

8.3.1 The REG/RAND paradigm

The REG/RAND paradigm still offers a viable alternative to oddball and related simpler repetition-based designs in the study of auditory scene analysis. It provides the benefit of manipulating predictability, both in the deterministic and entropic sense, independently of adaptation. However the EEG results in chapter 4 offer a warning: the two manipulations of predictability (repetition versus REG) show somewhat contradictory effects on brain responses. This implies that the repetition-based regularity paradigm is not generalisable, and is therefore not a good reductionist model of predictability to stand in for more naturalistic patterns. This has implications for the use of results from repetition paradigms to inform generalised predictive processing accounts (Garrido, Kilner, Kiebel, et al. 2009; Wacongne et al. 2012; Auksztulewicz and Friston 2016).

However, much more experimental work is required to delineate the realm of validity of the present regularity effects, before concluding that the regularity sustained response has a future beyond being a curious phenomenon.

For instance, if the sustained response increase truly represents the ‘state’ of predictability (or precision) the findings need to generalise to different physical properties, such as patterning expressed in timbre or loudness. The effect should also generalise to different presentation rates, to rule out the possibility that it is a rate-specific phenomenon such as the auditory steady-state response (Pantev et al. 1996). At least for the behavioural effects, the paradigm has already been applied to vision (Barascud 2015). It has also recently been deployed in animal models (Barczak et al. 2018), which are a powerful means of testing hypotheses regarding the activity of single neurons, inaccessible to scalp EEG.

8.3.1.1 An adequate model of predictability?

In some ways, the REG/RAND paradigm falls short of modelling predictable structure in a manner sufficiently resolved to map theoretical perceptual variables onto the brain response. For instance, predicted and predictive are indistinguishable in REG and RAND. These two factors are, however, subtly different. The distinction can be described as follows (Agres et al. 2017; see also Braem and Trapp 2017). Each tone represents the fulfilment (or violation) of a prediction based on the previous tone; i.e. it is more or less correctly predicted. In contrast, *predictive* refers to how informative each tone frequency is about the identity of the upcoming tone. In the present work (apart from the deviance paradigms) the predictive information and predicted information changed in perfect step. Each tone in REG, assuming perfect learning of the sequence, was both completely predictive and completely predicted. In RAND, each tone was equally unpredicted and unpredictive.

More specifically, in terms of the ramifications of the present paradigm for generalised predictive coding, this duality of predictive and predicted quantities lies at the core of failure to disambiguate between prediction activity and precision-weighting accounts of the increased responses observed throughout the EEG experiments. This represents a missed opportunity in studying how the brain tracks regularity, makes use of predicable structure and updates its beliefs in the face of new evidence during auditory scene analysis.

8.3.1.2 Extremes of predictability miss the point: the case of complexity

The most important comparison used for this thesis was that of REG versus RAND. This represents the extremes of predictability and unpredictability, respectively. The benefit of this manipulation was that the binary contrast allows for greater statistical power than, say, a parametric modulation, which was particularly crucial given the level of noise in the data (Lipsey 1990).

However (as argued by Bach and Dolan 2012; Hasson 2016), contrasting regular with random in a binary contrast may not be the most effective design to detect uncertainty-tracking mechanisms in the brain. The shape of the relationship between sequential predictability and other quantities of interest is not necessarily monotonic (Shiner et al. 1999). According to one view, complexity represents the opposite of compressibility, and as such a randomly-generated sequence represents high complexity as it requires many bits of information to perfectly encode the sequence (Shannon 1948). However, to model the generative process giving rise to the stimulus, complexity follows a non-monotonic, inverted-U relationship as a function of the degree of entropy in the stimulus (Crutchfield 2012). REG and RAND are both minimally complex by such a measure, as the underlying generative processes are simple to describe in each case.

The relationship between stimulus disorder and neural or behavioural responses is not necessarily monotonic. Such U-shaped responses to uncertainty have been observed in the brain and behaviour (Tobia et al. 2012; Nastase et al. 2015; Hasson 2016). Braem et al. (2017) show that participants' subjective preferences correlate with how predictive, not how predicted, a visual stimulus was in an implicit learning paradigm. Infants preferentially attend to auditory information which is of intermediate predictability relative to both more and less predictable stimuli (Kidd et al. 2014). In music, melodies also represent this intermediate level of predictability (Patel 2003), and it is this which makes music rewarding (Gebauer et al. 2012; Zatorre and Salimpoor 2013).

An alternative definition of salience, as distinguished from the tendency to draw exogenous attention (see Chapter 3), is related to the ability to resolve uncertainty (Friston et al. 2012). The epistemic value of a stimulus is low for both high and low predictability. Salient stimuli are those which are expected to minimise uncertainty (Parr and Friston 2017). In REG, uncertainty is rapidly resolved on formation of the auditory object. In RAND, no uncertainty is resolved by continuing to listen to the sequence. Therefore, both are in some way ‘uninteresting’ to the brain, notwithstanding the fact that regularity was mostly manipulated outside the context of any behavioural goals in the experiments presented in this thesis. Therefore, future work may explore sequences incorporating intermediate points in complexity-entropy space, which can implicitly vary the predictive information and uncertainty-resolving properties associated with the individual tones.

8.3.2 EEG data quality

Tracking the system of regularity extraction with the types of EEG paradigms used here may have a limit. A recurring problem throughout this thesis was that of relatively noisy EEG data. The slow drift was particularly problematic as this was hard to distinguish from the slow sustained response changes, and as discussed in Chapter 7, there is evidence that the denoising techniques did not fully remove this effect.

The group-level results obtained are in some cases marginally or intermittently significant, for instance compare the REG10 versus RAND10 effects in Chapters 3, 5, and 6. The conclusions are also drawn on the basis of trial-averaged data for each subject. The very approach used to help de-noise the data, namely averaging over trials, means that a huge amount of information in the EEG recording pertaining to the subject’s internal state is lost. Studying the effect of predictability in a more nuanced manner, such as manipulations of complexity on a fine-grained continuum, would benefit from single-trial data modelled using regressors for information-theoretic quantities (for instance Crosse et al. 2016; e.g. as used by Sedley et al. 2016; Broderick et al. 2017).

MEG may be a preferable imaging modality for further research on slow potentials, as well as further work using intracranial recordings and fMRI. Single-cell and multi-unit recordings in animal models are needed to establish responses to individual tones, and how this is influenced by contextual regularity manipulations, rather than inferring this from ensemble-level responses. Better spatial resolution of source activity would allow testing hypotheses of how representations of stimuli in different areas interact, using for example dynamic causal modelling (e.g. Phillips et al. 2015). This could be achieved using individual anatomical MRI scans in conjunction with EEG or MEG.

8.4 Future Work

Several possible continuations of the experiments in this thesis are suggested below.

8.4.1 Complexity and predictability

To address the potential criticism highlighted in §8.3.1, future work could use sequences representing intermediate states of predictability. An alternative view of the sequences used in this thesis is that they are Markov chains generated from a transition matrix instantiating the rules for determining the next note based on the previous (note the ideal observer model was based on representing the transition probabilities in the sequence). In the proposed experiment, sequences would be generated tone-by-tone from four types of transition matrix: deterministic (equivalent to REG); random (uniform transition matrix, equivalent to RAND); an interpolation between the transition matrices for REG and RAND; and finally a 'biased' condition where some rows of the transition matrix represent a flatter probability distribution (akin to RAND) and others more peaked, so that some notes are rendered more informative in terms of reducing uncertainty about the next note, and others are uninformative. Measures of overall sequence entropy and complexity could be derived from statistical models. Tone-wise measures of surprisal and predictive information could also be estimated (Abdallah and Plumbley 2010; Agres et al. 2017). Both these measures could be used as regressors to disentangle prediction, precision and surprise-related responses. The sequence-level measures could be used to test for monotonic parametric relationships between sequence-level entropy (interpreted as precision) and the magnitude of the sustained and offset responses, consistent with the hypothesis that these brain responses track precision. Furthermore, as discussed by Hasson (2016), it is suspected that an inverted-U relationship between sequence entropy and neural responses will be detectible, suggesting sensitivity to the complexity of the inferred generative source of the sequence. At the level of individual tones, the degree of surprise, and the strength of predictive information available regarding the next tone, could be quantified and related to prediction error and prediction respectively. It is likely that such sequences would have to be slower than the current 50-ms rate in order to disentangle these effects. A behavioural study could also be used, where subjects must respond to each tone in the sequence with a different button representing each tone frequency, and the reaction time could be used as a measure of surprise (Agres et al. 2017).

8.4.2 Layer-specific evoked responses

In order to more precisely test the proposed neural implementation for generalised predictive coding (Bastos et al. 2012), layer-specific responses to REG and RAND stimuli could be measured with high-precision MEG (Troebinger et al. 2014; Bonaiuto et al. 2017; Meyer et al. 2017). Signal originating from superficial layers should correspond to precision-weighted prediction error. Conversely, signal in deep layers should correlate with predictions themselves. Using tone-pip sequences with an altered experimental design such as discussed in the previous section, predictability of the sequence could be manipulated dissociably from prediction errors. Precision, prediction and prediction error could be quantitatively specified from a perceptual model, (Pearce et al. 2010; e.g. Mathys 2014) and used as a regressor to quantify the contribution of each to the evoked response localised to deep and superficial layers.

8.4.3 Top-down attentional set

The present results reveal that spectrotemporal patterns are learnt and encoded automatically, and generally outside of conscious awareness, which is consistent with findings from the auditory evoked response and statistical learning literature that even pre-attentive auditory scene analysis permits sophisticated recognition of regularities. So far, the EEG response to REG and RAND has always been measured in naïve, distracted listeners; whether the sustained evoked response difference would endure when sequences are attended remains to be seen. Attention has previously been shown to enhance or even reverse the difference in response magnitude between predicted and unpredicted stimuli (Rahnev, et al. 2012; Chennu et al. 2013; Auksztulewicz and Friston 2015; Hsu et al. 2018), though most studies compare show attentional manipulations alter unattended expectation suppression, rather than starting from a baseline of unattended expectation enhancement as seen here.

Attention could be focussed on the stimuli in two ways. First, attentional set could be manipulated to focus on the auditory modality as a whole, but without specifically targeting the predictable structure of the sequence, such that regularity is task-irrelevant. For example, a possible task could be counting noise bursts which are equally easily detected in REG and RAND, then reporting the count at the end of the block, thus maintaining attention on sound without confounding with task difficulty (see §6.7). Alternatively, the experimental design could render the detection of predictability explicitly relevant for the task, such as producing a likely continuation of the sequence or detecting a deviant (see §5.6, §6.6).

8.4.4 Generalisability

The paradigm could be tested using a wider variety of stimuli beyond sinusoidal tone pip streams. This could be wider-band complex sounds or sequences with slower dynamics. Is the increased sustained response to REG specific to the particular stimuli used here or is it a robust marker of sequence predictability? If the latter, the REG/RAND paradigm could provide a powerful tool for manipulating sensory precision, and investigating its effects on widespread brain and behavioural processes.

References

Abdallah S, Plumbley M. 2010. Information dynamics: patterns of expectation and surprise in the perception of music. Connection Science. 21:89–117.

Aghamolaei M, Zarnowiec K, Grimm S, Escera C. 2016. Functional dissociation between regularity encoding and deviance detection along the auditory hierarchy. Eur J Neurosci. 43:529–535.

Agres K, Abdallah S, Pearce M. 2017. Information-Theoretic Properties of Auditory Sequences Dynamically Influence Expectation and Memory. Cogn Sci. 21:89–34.

Agus TR, Thorpe SJ, Pressnitzer D. 2010. Rapid Formation of Robust Auditory Memories: Insights from Noise. NEURON. 66:610–618.

Ahlfors SP, Han J, Lin F-H, Witzel T, Belliveau JW, Hämäläinen MS, Halgren E. 2009. Cancellation of EEG and MEG signals generated by extended and distributed sources. Hum Brain Mapp. 22:NA–NA.

Aitchison L, Lengyel M. 2017. With or without you: predictive coding and Bayesian inference in the brain. Current Opinion in Neurobiology. 46:219–227.

Alamia A, Zenon A. 2016. Statistical Regularities Attract Attention when Task-Relevant. Front Hum Neurosci. 10:185.

Alho K. 1995. Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. Ear and Hearing. 16:38–51.

Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L. 2010. Stimulus predictability reduces responses in primary visual cortex. J Neurosci. 30:2960–2966.

Amenedo E, Escera C. 2000. The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. Eur J Neurosci. 12:2570–2574.

Andersen LM, Lundqvist D. 2019. Somatosensory responses to nothing: An MEG study of expectations during omission of tactile stimulations. NeuroImage. 184:78–89.

Andreou L-V, Griffiths TD, Chait M. 2015. Sensitivity to the temporal structure of rapid sound sequences — An MEG study. NeuroImage. 110:194–204.

Andreou L-V, Kashino M, Chait M. 2011. The role of temporal regularity in auditory segregation. Hearing Research. 280:228–235.

Andreou LV, Chait M. 2010. Effect of temporal regularity on segregation. In: (Proceedings) British Society for Audiology (2010).

Antunes FM, Nelken I, Covey E, Malmierca MS. 2010. Stimulus-Specific Adaptation in the Auditory Thalamus of the Anesthetized Rat. PLoS ONE. 5:e14071.

Attias H, Schreiner CE. 1997. Temporal low-order statistics of natural sounds. In:. Presented at the Advances in Neural Information Processing Systems. p. 27–33.

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

Auksztulewicz R, Friston K. 2015. Attentional Enhancement of Auditory Mismatch Responses: a DCM/MEG Study. Cerebral Cortex. 1–11.

Auksztulewicz R, Friston K. 2016. Repetition suppression and its contextual determinants in predictive coding. Cortex. 80:125–140.

Ayala YA, Pérez-González D, Malmierca MS. 2016. Stimulus-specific adaptation in the inferior colliculus: The role of excitatory, inhibitory and modulatory inputs. Biological Psychology. 116:10–22.

Bach DR, Dolan RJ. 2012. Knowing how much you don't know: a neural organization of uncertainty estimates. 1–15.

Bader M, Schröger E, Grimm S. 2017. How regularity representations of short sound patterns that are based on relative or absolute pitch information establish over time: An EEG study. PLoS ONE. 12:e0176981–23.

Baldeweg T. 2006. Repetition effects to sounds: evidence for predictive coding in the auditory system. Trends in Cognitive Sciences. 10:93–94.

Baldeweg T. 2008. ERP Repetition Effects and Mismatch Negativity Generation. Journal of Psychophysiology. 21:204–213.

Baldeweg T, Klugman A, Gruzelier J, Hirsch SR. 2004. Mismatch negativity potentials and cognitive impairment in schizophrenia. Schizophrenia Research. 69:203–217.

Barakat BK, Seitz AR, Shams L. 2013. The effect of statistical learning on internal stimulus representations: Predictable items are enhanced even when not predicted. Cognition. 129:205–211.

Barascud N. 2015. Auditory Pattern Detection. Doctoral thesis, UCL (University College London).

Barascud N, Griffiths TD, McAlpine D, Chait M. 2014. “Change Deafness” Arising from Inter-feature Masking within a Single Auditory Object. Journal of Cognitive Neuroscience. 26:514–528.

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

Barczak A, O'Connell MN, McGinnis T, Ross D, Mowery T, Falchier A, Lakatos P. 2018. Top-down, contextual entrainment of neuronal oscillations in the auditory thalamocortical circuit. Proc Natl Acad Sci USA. 115:E7605–E7614.

Barnes R, Jones MR. 2000. Expectancy, Attention, and Time. Cognitive Psychology. 41:254–311.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. 2012. Canonical Microcircuits for Predictive Coding. NEURON. 76:695–711.

Beagley HA, Knight JJ. 1967. Changes in Auditory Evoked Response with Intensity. The Journal of Laryngology & Otology. 81:861–873.

Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L. 2009. Neural signature of the conscious processing of auditory regularities. Proc Natl Acad Sci USA. 106:1672–1677.

Bell AH, Summerfield C, Morin EL, Malecek NJ, Ungerleider LG. 2016. Encoding of Stimulus Probability in Macaque Inferior Temporal Cortex. Current Biology. 26:2280–2290.

Bendixen A. 2014. Predictability effects in auditory scene analysis: a review. Front Neurosci. 8:60–60.

Bendixen A, Bőhm TM, Szalárdy O, Mill R, Denham SL, Winkler I. 2013. Different roles of similarity and predictability in auditory stream segregation. Learning & Perception. 5:37–54.

Bendixen A, Denham SL, Gyimesi K, Winkler I. 2010. Regular patterns stabilize auditory streams. J Acoust Soc Am. 128:3658–3666.

Bendixen A, SanMiguel I, Schröger E. 2012. Early electrophysiological indicators for predictive processing in audition: A review. 83:120–131.

Bendixen A, Schröger E, Ritter W, Winkler I. 2012. Regularity extraction from non-adjacent sounds. Front Psychol. 3:143.

Bendixen A, Schröger E, Winkler I. 2009. I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. J Neurosci. 29:8447–8451.

Berwick RC, Okanoya K, Beckers GJL, Bolhuis JJ. 2011. Songs to syntax: the linguistics of birdsong. Trends in Cognitive Sciences. 15:113–121.

Bestmann S, Harrison LM, Blankenburg F, Mars RB, Haggard P, Friston KJ, Rothwell JC. 2008. Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. Current Biology. 18:775–780.

Bianco R, Novembre G, Keller PE, Kim S-G, Scharf F, Friederici AD, Villringer A, Sammler D. 2016. Neural networks for harmonic structure in music perception and action. NeuroImage. 142:454–464.

Billings JCW, Medda A, Shakil S, Shen X, Kashyap A, Chen S, Abbas A, Zhang X, Nezafati M, Pan W-J, Berman GJ, Keilholz SD. 2017. Instantaneous brain dynamics mapped to a continuous state space. NeuroImage. 162:344–352.

Blank H, Davis MH. 2016. Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. PLoS Biol. 14:e1002577–32.

Bonaiuto JJ, Rossiter HE, Meyer SS, Adams N, Little S, Callaghan MF, Dick F, Bestmann S, Barnes GR. 2017. Non-invasive laminar inference with MEG:

Comparison of methods and source inversion algorithms. NeuroImage. 167:372–383.

Bradbury JW, Vehrencamp SL. 1998. Principles of animal communication.

Braem S, Trapp S. 2017. Humans show a higher preference for stimuli that are predictive relative to those that are predictable. Psychol Res. 1–7.

Bregman AS. 1990. Auditory scene analysis.

Bregman AS, Campbell J. 1971. Primary Auditory Stream Segregation and Perception of Order in Rapid Sequences of Tones. Journal of Experimental Psychology. 89:244–&.

Briley PM, Krumbholz K. 2013. The specificity of stimulus-specific adaptation in human auditory cortex increases with repeated exposure to the adapting stimulus. Journal of Neurophysiology. 110:2679–2688.

Broadbent DE. 1952. Listening to one of two synchronous messages. Journal of Experimental Psychology. 44:51–55.

Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, Lalor EC. 2017. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. 1–18.

Budd TW, Barry RJ, Gordon E, Rennie C, Michie PT. 1998. Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. International Journal of Psychophysiology. 31:51–68.

Bullock TH, Karamürsel S, Achimowicz JZ, McClune MC, Başar-Eroglu C. 1994. Dynamic properties of human visual evoked and omitted stimulus potentials. Electroencephalogr Clin Neurophysiol. 91:42–53.

Calderone DJ, Lakatos P, Butler PD, Castellanos FX. 2014. Entrainment of neural oscillations as a modifiable substrate of attention. Trends in Cognitive Sciences. 18:300–309.

Carlyon RP. 2004. How the brain separates sounds. Trends in Cognitive Sciences. 8:465–471.

Cervantes Constantino F, Pinggera L, Paranamana S, Kashino M, Chait M. 2012. Detection of Appearing and Disappearing Objects in Complex Acoustic Scenes. PLoS ONE. 7:e46167–13.

Chait M, Poeppel D, de Cheveigné A, Simon JZ. 2007. Processing Asymmetry of Transitions between Order and Disorder in Human Auditory Cortex. Journal of Neuroscience. 27:5207–5214.

Chait M, Poeppel D, de Cheveigné A, Simon JZ. 2005. Human auditory cortical processing of changes in interaural correlation. J Neurosci. 25:8518–8527.

Chait M, Poeppel D, de Cheveigné A, Simon JZ. 2007. Processing asymmetry of transitions between order and disorder in human auditory cortex. J Neurosci. 27:5207–5214.

Chait M, Poeppel D, Simon JZ. 2007. Stimulus context affects auditory cortical responses to changes in interaural correlation. Journal of Neurophysiology. 98:224–231.

Chait M, Poeppel D, Simon JZ. 2008. Auditory temporal edge detection in human auditory cortex. Brain Research. 1213:78–90.

Chait M, Ruff CC, Griffiths TD, McAlpine D. 2012. Cortical responses to changes in acoustic regularity are differentially modulated by attentional load. NeuroImage. 59:1932–1941.

Chalk M, Seitz AR, Seriès P. 2010. Rapidly learned stimulus expectations alter perception of motion. Journal of Vision. 10:2–2.

Chambers C, Akram S, Adam V, Pelofi C, Sahani M, Shamma S, Pressnitzer D. 2017. Prior context in audition informs binding and shapes simple features. Nature Communications. 8:15027.

Chen I-W, Helmchen F, Lütcke H. 2015. Specific Early and Late Oddball-Evoked Responses in Excitatory and Inhibitory Neurons of Mouse Auditory Cortex. J Neurosci. 35:12560–12573.

Chennu S, Bekinschtein TA. 2012. Arousal modulates auditory attention and awareness: insights from sleep, sedation, and disorders of consciousness. Front Psychol. 3:65.

Chennu S, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibanez A, Owen AM, Bekinschtein TA. 2013. Expectation and Attention in Hierarchical Auditory Prediction. Journal of Neuroscience. 33:11194–11205.

Chennu S, Noreika V, Gueorguiev D, Shtyrov Y, Bekinschtein TA, Henson R. 2016. Silent Expectations: Dynamic Causal Modeling of Cortical Prediction and Attention to Sounds That Weren't. Journal of Neuroscience. 36:8305–8316.

Cherry EC. 1953. Some Experiments on the Recognition of Speech, with One and with Two Ears. J Acoust Soc Am. 25:975–979.

Chouiter L, Tzovara A, Dieguez S, Annoni J-M, Magezi D, De Lucia M, Spierer L. 2015. Experience-based Auditory Predictions Modulate Brain Activity to Silence as Do Real Sounds. Journal of Cognitive Neuroscience. 27:1968–1980.

Christianson GB, Chait M, de Cheveigné A, Linden JF. 2014. Auditory evoked fields measured noninvasively with small-animal MEG reveal rapid repetition suppression in the guinea pig. Journal of Neurophysiology. 112:3053–3065.

Chun MM, Jiang Y. 1999. Top-Down Attentional Guidance Based on Implicit Learning of Visual Covariation. Psychological Science. 10:360–365.

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

Clark A. 2015. Surfing uncertainty: Prediction, action, and the embodied mind.

Conway CM, Christiansen MH. 2005a. Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. 31:24–39.

Conway CM, Christiansen MH. 2005b. Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition. 31:24–39.

Correa Á, Nobre AC. 2008. Neural modulation by regularity and passage of time. Journal of Neurophysiology. 100:1649–1655.

Costa-Faidella J, Baldeweg T, Grimm S, Escera C. 2011. Interactions between "what" and “when” in the auditory system: temporal predictability enhances repetition suppression. J Neurosci. 31:18590–18597.

Costa-Faidella J, Grimm S, Slabu L, Díaz-Santaella F, Escera C. 2010. Multiple time scales of adaptation in the auditory system as revealed by human evoked potentials. Psychophysiology. 48:774–783.

Courville AC, Daw ND, Touretzky DS. 2006. Bayesian theories of conditioning in a changing world. Trends in Cognitive Sciences. 10:294–300.

Cowan N, Winkler I, Teder W, Näätänen R. 1993. Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). Journal of Experimental Psychology: Learning, Memory, and Cognition. 19:909–921.

Crosse MJ, Di Liberto GM, Bednar A, Lalor EC. 2016. The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. Front Hum Neurosci. 10:3958–14.

Crutchfield JP. 2012. Between order and chaos. Nature Physics. 8:17–24.

Cusack R. 2005. The intraparietal sulcus and perceptual organization. Journal of Cognitive Neuroscience. 17:641–651.

Cusack R, Roberts B. 2004. Effects of differences in the pattern of amplitude envelopes across harmonics on auditory stream segregation. Hearing Research. 193:95–104.

Daikhin L, Ahissar M. 2012. Responses to deviants are modulated by subthreshold variability of the standard. Psychophysiology. 49:31–42.

Daikoku, T., 2018. Neurophysiological Markers of Statistical Learning in Music and Language: Hierarchy, Entropy, and Uncertainty. *Brain Sciences*, 8(6), pp.114–23.

Dale AM, Liu AK, Fischl BR, Buckner RL, Belliveau JW, Lewine JD, Halgren E. 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. NEURON. 26:55–67.

Daubechies I. 1992. Ten lectures on wavelets.

Dayan P, Hinton GE, Neal RM, Zemel RS. 1995. The Helmholtz Machine. http://dxdoiorg/101162/neco199575889. 7:889–904.

Dayan P, Kakade S, Montague PR. 2000. Learning and selective attention. Nat Neurosci. 3 Suppl:1218–1223.

Dayan P, Yu AJ. 2002. ACh, uncertainty, and cortical inference. Advances in neural information processing ….

de C Hamilton AF, Grafton ST. 2009. Repetition suppression for performed hand gestures revealed by fMRI. Hum Brain Mapp. 30:2898–2906.

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

de Cheveigné A, Nelken I. 2019. Filters: When, Why, and How (Not) to Use Them. NEURON. 102:280–293.

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

de Cheveigné A, Simon JZ. 2008. Denoising based on spatial filtering. Journal of Neuroscience Methods. 171:331–339.

de Lange FP, Heilbron M, Kok P. 2018. How Do Expectations Shape Perception? Trends in Cognitive Sciences. 22:764–779.

Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C. 2015. The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. NEURON. 88:2–19.

Deike S, Gaschler-Markefski B, Brechmann A, Scheich H. 2004. Auditory stream segregation relying on timbre involves left auditory cortex. Neuroreport. 15:1511–1514.

Delcomyn F. 1980. Neural basis of rhythmic behavior in animals. Science. 210:492–498.

Delorme A, Makeig S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods. 134:9–21.

Denham SL, Winkler I. 2006. The role of predictive models in the formation of auditory streams. Journal of Physiology - Paris. 100:154–170.

Denham SL, Winkler I. 2018. Predictive coding in auditory perception: challenges and unresolved questions. Eur J Neurosci. 66:610–610.

Desimone R. 1996. Neural mechanisms for visual memory and their role in attention. Proc Natl Acad Sci USA. 93:13494–13499.

Deutsch D. 1974. An auditory illusion. Nature. 251:307–309.

Devergie A, Grimault N, Tillmann B, Berthommier F. 2010. Effect of rhythmic attention on the segregation of interleaved melodies. The Journal of the Acoustical Society of America. 128:EL1–EL7.

Ding N, Melloni L, Yang A, Wang Y, Zhang W, Poeppel D. 2017. Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG). Front Hum Neurosci. 11:2095–2099.

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

Drake C, Botte M-C. 1993. Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. Percept Psychophys. 54:277–286.

Efron B, Tibshirani RJ. 1993. An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability, Vol. 57. New York and London: Chapman and Hall/CRC.

Egermann H, Pearce MT, Wiggins GA, McAdams S. 2013. Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. Cognitive, Affective, & Behavioral Neuroscience. 13:533–553.

Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA. 2009. Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes. NEURON. 61:317–329.

Eliades SJ, Wang X. 2008. Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature. 453:1102–1106.

Ellis RJ, Jones MR. 2010. Rhythmic context modulates foreperiod effects. Atten Percept Psychophys. 72:2274–2288.

Escera C, Alho K, Winkler I, Näätänen R. 1998. Neural Mechanisms of Involuntary Attention to Acoustic Novelty and Change. Journal of Cognitive Neuroscience. 10:590–604.

Escera C, Leung S, Grimm S. 2014. Deviance Detection Based on Regularity Encoding Along the Auditory Hierarchy: Electrophysiological Evidence in Humans. Brain Topogr. 27:527–538.

Escera C, Malmierca MS. 2014. The auditory novelty system: an attempt to integrate human and animal research. Psychophysiology. 51:111–123.

Fairhall AL, Lewen GD, Bialek W, de Ruyter Van Steveninck RR. 2001. Efficiency and ambiguity in an adaptive neural code. Nature. 412:787–792.

Falk R, Konold C. 1997. Making sense of randomness: Implicit encoding as a basis for judgment. Psychol Rev. 104:301–318.

Feldman H, Friston KJ. 2010. Attention, Uncertainty, and Free-Energy. Front Hum Neurosci. 4:1–23.

Feuerriegel D, Churches O, Coussens S, Keage HAD. 2017. Evidence for spatiotemporally distinct effects of image repetition and perceptual expectations as measured by event-related potentials. NeuroImage. 1–40.

Fiser J, Aslin RN. 2001. Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. Psychological Science. 12:499–504.

Fishman YI, Arezzo JC, Steinschneider M. 2004. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. The Journal of the Acoustical Society of America. 116:1656–1670.

Fishman YI, Reser DH, Arezzo JC, Steinschneider M. 2001. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hearing Research. 151:167–187.

Frey S, Kostopoulos P, Petrides M. 2004. Orbitofrontal contribution to auditory encoding. NeuroImage. 22:1384–1389.

Friedman D, Cycowicz YM, Gaeta H. 2001. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. Neuroscience and Biobehavioral Reviews. 25:355–373.

Friston K. 2005. A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biological Sciences. 360:815–836.

Friston K. 2008. Hierarchical Models in the Brain. PLoS Comput Biol. 4:e1000211–e1000224.

Friston K, Adams RA, Perrinet L, Breakspear M. 2012. Perceptions as hypotheses: saccades as experiments. Front Psychol. 3:151–151.

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

Friston K, Kilner J, Harrison L. 2006. A free energy principle for the brain. Journal of Physiology - Paris. 100:70–87.

Friston K, Rigoli F, Ognibene D, Mathys C. 2015. Active inference and epistemic value. Neuroscience.

Friston KJ, Daunizeau J, Kilner J, Kiebel SJ. 2010. Action and behavior: a free-energy formulation. Biol Cybern. 102:227–260.

Fuchs M, Kastner J, Wagner M, Hawes S, Ebersole JS. 2002. A standardized boundary element method volume conductor model. Clinical Neurophysiology. 113:702–712.

Furl N, Kumar S, Alter K, Durrant S, Shawe-Taylor J, Griffiths TD. 2011. Neural prediction of higher-order auditory sequence statistics. 54:2267–2277.

Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T, Kilner JM. 2008. The functional anatomy of the MMN: a DCM study of the roving paradigm. NeuroImage. 42:936–944.

Garrido MI, Kilner JM, Kiebel SJ, Friston KJ. 2007. Evoked brain responses are generated by feedback loops. Proc Natl Acad Sci USA. 104:20961–20966.

Garrido MI, Kilner JM, Kiebel SJ, Stephan KE, Baldeweg T, Friston KJ. 2009. Repetition suppression and plasticity in the human brain. NeuroImage. 48:269–279.

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

Garrido MI, Sahani M, Dolan RJ. 2013. Outlier Responses Reflect Sensitivity to Statistical Structure in the Human Brain. PLoS Comput Biol. 9:e1002999–10.

Gebauer L, Kringelbach ML, Vuust P. 2012. Ever-changing cycles of musical pleasure: The role of dopamine and anticipation. Psychomusicology: Music, Mind, and Brain. 22:152–167.

Gebhart AL, Newport EL, Aslin RN. 2009. Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. Psychonomic Bulletin & Review. 16:486–490.

Geiser E, Notter M, Gabrieli JDE. 2012. A corticostriatal neural system enhances auditory perception through temporal context processing. J Neurosci. 32:6177–6182.

Gelman A, Stern H. 2006. The Difference Between 'Significant' and “Not Significant” is not Itself Statistically Significant. The American Statistician. 60:328–331.

Gifford AM, Sperling MR, Sharan A, Gorniak RJ, Williams RB, Davis K, Kahana MJ, Cohen YE. 2018. Neuronal phase consistency tracks dynamic changes in acoustic spectral regularity. Eur J Neurosci. 812:23–20.

Godey B, Schwartz D, de Graaf JB, Chauvel P, Liégeois-Chauvel C. 2001. Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. Clinical Neurophysiology. 112:1850–1859.

Graf Estes K, Evans JL, Alibali MW, Saffran JR. 2007. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. Psychological Science. 18:254–260.

Grech R, Cassar T, Muscat J, Camilleri KP, Fabri SG, Zervakis M, Xanthopoulos P, Sakkalis V, Vanrumste B. 2008. Review on solving the inverse problem in EEG source analysis. J NeuroEngineering Rehabil. 5:25–33.

Green B, Jääskeläinen IP, Sams M, Rauschecker JP. 2018. Distinct brain areas process novel and repeating tone sequences. Brain Lang. 1–0.

Gregory RL. 1980. Perceptions as hypotheses. Philosophical Transactions of the Royal Society B: Biological Sciences. 290:181–197.

Grill-Spector K, Henson R, Martin A. 2006. Repetition and the brain: neural models of stimulus-specific effects. 10:14–23.

Grimault N, Bacon SP, Micheyl C. 2002. Auditory stream segregation on the basis of amplitude-modulation rate. The Journal of the Acoustical Society of America. 111:1340–1348.

Grimm S, Escera C, Nelken I. 2016. Early indices of deviance detection in humans and animal models. Biological Psychology. 116:23–27.

Grimm S, Escera C, Slabu L, Faidella JC. 2011. Electrophysiological evidence for the hierarchical organization of auditory change detection in the human brain. Psychophysiology. 48:377–384.

Grotheer M, Kovács G. 2015. The relationship between stimulus repetitions and fulfilled expectations. Neuropsychologia. 67:175–182.

Grotheer M, Kovács G. 2016. Can predictive coding explain repetition suppression? Cortex. 80:113–124.

Hadamard J. 1923. Lectures on Cauchy's problem in linear partial differential equations.

Haenschel C, Vernon DJ, Dwivedi P, Gruzelier JH, Baldeweg T. 2005. Event-related brain potential correlates of human auditory sensory memory-trace formation. 25:10494–10501.

Halgren E, Sherfey J, Irimia A, Dale AM, Marinkovic K. 2010. Sequential temporo-fronto-temporal activation during monitoring of the auditory environment for temporal patterns. Hum Brain Mapp. 32:1260–1276.

Hansen NC, Pearce MT. 2014. Predictive uncertainty in auditory sequence processing. Front Psychol. 5:1052.

Hari R, Pelizzone M, Mäkelä JP, Hällström J, Leinonen L, Lounasmaa OV. 1987. Neuromagnetic Responses of the Human Auditory Cortex to On- and Offsets of Noise Bursts: Réponses neuromagnétiques du cortex auditif humain à l"établissement et l"extinction de salves de bruit. Int J Audiol. 26:31–43.

Hasson U. 2016. The neurobiology of uncertainty: implications for statistical learning. Philosophical Transactions of the Royal Society B: Biological Sciences. 372:20160048–10.

Heilbron M, Chait M. 2017. Great expectations: Is there evidence for predictive coding in auditory cortex? Neuroscience.

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

Helmholtz H, Ellis AJ. 1875. On the sensation of sound in general.

Henson RN, Mouchlianitis E, Matthews WJ, Kouider S. 2008. Electrophysiological correlates of masked face priming. NeuroImage. 40:884–895.

Herrmann B, Henry MJ, Fromboluti EK, McAuley JD, Obleser J. 2015. Statistical context shapes stimulus-specific adaptation in human auditory cortex. Journal of Neurophysiology. 113:2582–2591.

Herrmann B, Henry MJ, Obleser J. 2013. Frequency-specific adaptation in human auditory cortex depends on the spectral variance in the acoustic stimulation. Journal of Neurophysiology. 109:2086–2096.

Herrmann B, Johnsrude IS. 2018. Neural Signatures of the Processing of Temporal Patterns in Sound. Journal of Neuroscience. 38:5466–5477.

Hillyard SA, Picton TW. 1978. On and off components in the auditory evoked potential. Percept Psychophys. 24:391–398.

Hohwy J. 2013. The predictive mind. Oxford University Press.

Hsu Y-F, Hämäläinen JA, Waszak F. 2014. Both attention and prediction are necessary for adaptive neuronal tuning in sensory processing. Front Hum Neurosci. 8:14608.

Hsu Y-F, Hämäläinen JA, Waszak F. 2018. The processing of mispredicted and unpredicted sensory inputs interact differently with attention. Neuropsychologia. 111:85–91.

Huang N, Elhilali M. 2017. Auditory salience using natural soundscapes. The Journal of the Acoustical Society of America. 141:2163–2176.

Hughes HC, Darcey TM, Barkan HI, Williamson PD, Roberts DW, Aslin CH. 2001. Responses of human auditory association cortex to the omission of an expected acoustic event. NeuroImage. 13:1073–1089.

Huigen E, Peper A, Grimbergen CA. 2002. Investigation into the origin of the noise of surface electrodes. Med Biol Eng Comput. 40:332–338.

Itti L, Baldi P. 2009. Bayesian surprise attracts human attention. Vision Research. 49:1295–1306.

Itti L, Koch C. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research. 40:1489–1506.

Itti L, Koch C. 2001. Feature combination strategies for saliency-based visual attention systems. J Electron Imaging. 10:161–169.

Janata P, Grafton ST. 2003. Swinging in the brain: Shared neural substrates for behaviors related to sequencing and music. Nat Neurosci. 6:682–687.

Jaramillo S, Zador AM. 2011. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. Nat Neurosci. 14:246–251.

Jaunmahomed Z, Chait M. 2012. The timing of change detection and change perception in complex acoustic scenes. Front Psychol. 3:396.

Jääskeläinen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levänen S, Lin F-H, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW. 2004. Human posterior auditory cortex gates novel sounds to consciousness. Proc Natl Acad Sci USA. 101:6809–6814.

Jensen O, Lisman JE. 2005. Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer. Trends in Neurosciences. 28:67–72.

Jones MR, Johnston HM, Puente J. 2006. Effects of auditory pattern structure on anticipatory and reactive attending. Cognitive Psychology. 53:59–96.

Jones MR, Moynihan H, MacKenzie N, Puente J. 2002. Temporal Aspects of Stimulus-Driven Attending in Dynamic Arrays. Psychological Science. 13:313–319.

Kanai R, Komura Y, Shipp S, Friston K. 2015. Cerebral hierarchies: predictive processing, precision and the pulvinar. Philosophical Transactions of the Royal Society B: Biological Sciences. 370:20140169–20140169.

Kappenman ES, Luck SJ. 2010. The effects of electrode impedance on data quality and statistical significance in ERP recordings. Psychophysiology. 47:888–904.

Khouri L, Nelken I. 2015. Detecting the unexpected. Current Opinion in Neurobiology. 35:142–147.

Kidd C, Piantadosi ST, Aslin RN. 2014. The Goldilocks Effect in Infant Auditory Attention. Child Dev. 19:n/a–n/a.

Kidd G, Mason CR, Deliwala PS, Woods WS, Colburn HS. 1994. Reducing informational masking by sound segregation. The Journal of the Acoustical Society of America. 95:3475–3480.

Kidd G, Mason CR, Richards VM. 2003. Multiple bursts, multiple looks, and stream coherence in the release from informational masking. The Journal of the Acoustical Society of America. 114:2835–12.

Kleiner M, Brainard D, Pelli D, Ingling A. 2007. What's new in Psychtoolbox-3. In:. Presented at the Perception 36 ECVP Abstract Supplement.

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

Koelsch S, Busch T, Jentschke S, Rohrmeier M. 2016. Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. Sci Rep. 6:19741.

Koelsch S, Gunter T, Friederici AD, Schröger E. 2000. Brain indices of music processing: “nonmusicians” are musical. Journal of Cognitive Neuroscience. 12:520–541.

Koffka K. 1935. Principles of Gestalt psychology. Oxford, England: Harcourt, Brace SN .

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

Kok P, Rahnev D, Jehee JFM, Lau HC, de Lange FP. 2012. Attention Reverses the Effect of Prediction in Silencing Sensory Signals. Cerebral Cortex. 22:2197–2206.

Kumar S, Bonnici HM, Teki S, Agus TR, Pressnitzer D, Maguire EA, Griffiths TD. 2014. Representations of specific acoustic patterns in the auditory cortex and hippocampus. Proceedings of the Royal Society B: Biological Sciences. 281:20141000–20141000.

Lachaux JP, Rodriguez E, Martinerie J, Varela FJ. 1999. Measuring phase synchrony in brain signals. Hum Brain Mapp. 8:194–208.

Lange K. 2013. The ups and downs of temporal orienting: a review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. Front Hum Neurosci. 7:263.

Leaver AM, Van Lare J, Zielinski B, Halpern AR, Rauschecker JP. 2009. Brain activation during anticipation of sound sequences. J Neurosci. 29:2477–2485.

Lecaignard F, Bertrand O, Gimenez G, Mattout J, Caclin A. 2015. Implicit learning of predictable sound sequences modulates human brain responses at different levels of the auditory hierarchy. Front Hum Neurosci. 9:212–214.

León-Cabrera P, Flores A, Rodriguez-Fornells A, Morís J. 2019. Ahead of time: Early sentence slow cortical modulations associated to semantic prediction. NeuroImage. 189:192–201.

Lewicki, M.S., 2002. Efficient coding of natural sounds. *Nature Neuroscience*, 5(4), pp.356–363.

Liégeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P. 1994. Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. Electroencephalogr Clin Neurophysiol. 92:204–214.

Lipsey MW. 1990. Design sensitivity.

López JD, Litvak V, Espinosa JJ, Friston K, Barnes GR. 2014. Algorithmic procedures for Bayesian MEG/EEG source reconstruction in SPM. NeuroImage. 84:476–487.

Luck SJ. 2005. An introduction to the event-related potential technique, vol. 1.

Mackintosh NJ. 1975. A theory of attention: Variations in the associability of stimuli with reinforcement. Psychol Rev. 82:276–298.

Maess B, Koelsch S, Gunter TC, Friederici AD. 2001. Musical syntax is processed in Broca's area: an MEG study. Nat Neurosci. 4:540–545.

Maess B, Schröger E, Widmann A. 2016. High-pass filters and baseline correction in M/EEG analysis-continued discussion. Journal of Neuroscience Methods. 266:171–172.

Makieg S, Jung TP, Bell AJ, Ghahremani D, Sejnowski TJ. 1997. Blind separation of auditory event-related brain responses into independent components. Proc Natl Acad Sci USA. 94:10979–10984.

Malmierca MS, Cristaudo S, Pérez-González D, Covey E. 2009. Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. J Neurosci. 29:5483–5493.

Malmierca MS, Niño-Aguillón BE, Nieto-Diego J, Porteros Á, Pérez-González D, Escera C. 2019. Pattern-sensitive neurons reveal encoding of complex auditory regularities in the rat inferior colliculus. NeuroImage. 184:889–900.

Malmierca MS, Sanchez-Vives MV, Escera C, Bendixen A. 2014. Neuronal adaptation, novelty detection and regularity encoding in audition. Frontiers in Systems Neuroscience. 8:111.

Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods. 164:177–190.

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

Martin BA, Boothroyd A. 2000. Cortical, auditory, evoked potentials in response to changes of spectrum and amplitude. The Journal of the Acoustical Society of America. 107:2155–2161.

Masson R, Bidet-Caulet A. 2019. Fronto-central P3a to distracting sounds: An index of their arousing properties. 185:164–180.

Masutomi K, Barascud N, Kashino M, McDermott JH, Chait M. 2016. Sound segregation via embedded repetition is robust to inattention. Journal of Experimental Psychology: Human Perception and Performance. 42:386–400.

Mates J, Müller U, Radil T, Pöppel E. 1994. Temporal integration in sensorimotor synchronization. Journal of Cognitive Neuroscience. 6:332–340.

Mathys C. 2011. A Bayesian foundation for individual learning under uncertainty. 1–20.

Mathys CD. 2014. Uncertainty in perception and the Hierarchical Gaussian Filter. 1–24.

May P, Tiitinen H. 2001. Human cortical processing of auditory events over time. Neuroreport. 12:573–577.

May PJC, Tiitinen H. 2010. Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Psychophysiology. 47:66–122.

McDermott JH. 2009. The cocktail party problem. Curr Biol. 19:R1024–R1027.

McDermott JH, Schemitsch M, Simoncelli EP. 2013. Summary statistics in auditory perception. Nat Neurosci. 16:493–498.

McDermott JH, Wrobleski D, Oxenham AJ. 2011. Recovering sound sources from embedded repetition. Proc Natl Acad Sci USA. 108:1188–1193.

Meijs EL, Klaassen FH, Bokeria L, van Gaal S, de Lange FP. 2018. Cue predictability does not modulate bottom-up attentional capture. R Soc open sci. 5:180524.

Meyer SS, Bonaiuto J, Lim M, Rossiter H, Waters S, Bradbury D, Bestmann S, Brookes M, Callaghan MF, Weiskopf N, Barnes GR. 2017. Flexible head-casts for high spatial precision MEG. Journal of Neuroscience Methods. 276:38–45.

Meyer T, Olson CR. 2011. Statistical learning of visual transitions in monkey inferotemporal cortex. Proc Natl Acad Sci USA. 108:19401–19406.

Milne AE, Wilson B, Christiansen MH. 2018. Structured sequence learning across sensory modalities in humans and nonhuman primates. Current Opinion in Behavioral Sciences. 21:39–48.

Molloy K, Griffiths TD, Chait M, Lavie N. 2015. Inattentional Deafness: Visual Load Leads to Time-Specific Suppression of Auditory Evoked Responses. J Neurosci. 35:16046–16054.

Moran RJ, Campo P, Symmonds M, Stephan KE, Dolan RJ, Friston KJ. 2013. Free energy, precision and learning: the role of cholinergic neuromodulation. J Neurosci. 33:8227–8236.

Murakami S, Okada Y. 2006. Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. J Physiol. 575:925–936.

Murray MM, Brunet D, Michel CM. 2008. Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr. 20:249–264.

Mustovic H, Scheffler K, Di Salle F, Esposito F, Neuhoff JG, Hennig J, Seifritz E. 2003. Temporal integration of sequential auditory events: silent period in sound pattern activates human planum temporale. NeuroImage. 20:429–434.

Nastase S, Iacovella V, Hasson U. 2014. Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. Hum Brain Mapp. 35:1111–1128.

Nastase SA, Iacovella V, Ben Davis, Hasson U. 2015. Connectivity in the human brain dissociates entropy and complexity of auditory inputs. NeuroImage. 108:292–300.

Natan RG, Briguglio JJ, Mwilambwe-Tshilobo L, Jones SI, Aizenberg M, Goldberg EM, Geffen MN. 2015. Complementary control of sensory adaptation by two types of cortical interneurons. Elife. 4:7359.

Näätänen R, Gaillard AW, Mäntysalo S. 1978. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol (Amst). 42:313–329.

Näätänen R, Paavilainen P, Rinne T, Alho K. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clinical Neurophysiology. 118:2544–2590.

Näätänen R, Pakarinen S, Rinne T, Takegata R. 2004. The mismatch negativity (MMN): towards the optimal paradigm. Clinical Neurophysiology. 115:140–144.

Näätänen R, Schröger E, Karakas S, Tervaniemi M, Paavilainen P. 1993. Development of a memory trace for a complex sound in the human brain. Neuroreport. 4:503–506.

Nelken I. 2014. Stimulus-specific adaptation and deviance detection in the auditory system: experiments and models. Biol Cybern. 108:655–663.

Nelken, I., Rotman, Y. & Bar Yosef, O., 1999. Responses of auditory-cortex neurons to structural features of natural sounds. *Nature*, 397(6715), pp.154–157.

Nobre AC, Correa A, Coull JT. 2007. The hazards of time. Current Opinion in Neurobiology. 17:465–470.

Nordby H, Roth WT, Pfefferbaum A. 1988. Event-related potentials to breaks in sequences of alternating pitches or interstimulus intervals. Psychophysiology. 25:262–268.

Omigie D, Pearce MT, Williamson VJ, Stewart L. 2013. Electrophysiological correlates of melodic processing in congenital amusia. Neuropsychologia. 51:1749–1762.

Ono K, Altmann CF, Matsuhashi M, Mima T, Fukuyama H. 2015. Neural correlates of perceptual grouping effects in the processing of sound omission by musicians and nonmusicians. Hearing Research. 319:25–31.

Oostenveld R, Fries P, Maris E, Schoffelen J-M. 2010. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive

Electrophysiological Data. Computational Intelligence and Neuroscience. 2011:1–9.

Opitz B, Friederici AD. 2007. Neural basis of processing sequential and hierarchical syntactic structures. Hum Brain Mapp. 28:585–592.

Opitz B, Rinne T, Mecklinger A, Cramon von DY, Schröger E. 2002. Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. NeuroImage. 15:167–174.

Overath T, Cusack R, Kumar S, Kriegstein von K, Warren JD, Grube M, Carlyon RP, Griffiths TD. 2007. An Information Theoretic Characterisation of Auditory Encoding. PLoS Biol. 5:e288.

Paavilainen P. 2013. The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: a review. Int J Psychophysiol. 88:109–123.

Pantev C, Elbert T, Makieg S, Hampson S, Eulitz C, Hoke M. 1993. Relationship of Transient and Steady-State Auditory-Evoked Fields. Electroencephalogr Clin Neurophysiol. 88:389–396.

Pantev C, Roberts LE, Elbert T, Ross B, Wienbruch C. 1996. Tonotopic organization of the sources of human auditory steady-state responses. Hearing Research. 101:62–74.

Parr T, Friston KJ. 2017. Working memory, attention, and salience in active inference. Sci Rep. 1–21.

Parras GG, Nieto-Diego J, Carbajal GV, Valdés-Baizabal C, Escera C, Malmierca MS. 2017. Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. Nature Communications. 8:2148.

Patel AD. 2003. A new approach to the cognitive neuroscience of melody. The cognitive neuroscience of music.

Patel M, Chait M. 2011. Retroactive adjustment of perceived time. Cognition. 119:125–130.

Payzan-LeNestour E, Dunne S, Bossaerts P, O’Doherty JP. 2013. The Neural Representation of Unexpected Uncertainty during Value-Based Decision Making. NEURON. 79:191–201.

Pearce JM, Hall G. 1980. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev. 87:532–552.

Pearce JM, Mackintosh NJ. 2010. Two Theories of Attention: A Review and Possible Integration. In: Mitchell CJ, Le Pelley ME editors. Focusing attention on sound.

Pearce MT. 2005. The construction and evaluation of statistical models of melodic structure in music perception and composition. Doctoral thesis, QMUL (Queen Mary University London).

Pearce MT, Ruiz MH, Kapasi S, Wiggins GA, Bhattacharya J. 2010. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. NeuroImage. 50:302–313.

Petit J-P, Midgley KJ, Holcomb PJ, Grainger J. 2006. On the time course of letter perception: a masked priming ERP investigation. Psychonomic Bulletin & Review. 13:674–681.

Phillips DP, Hall SE, Boehnke SE. 2002. Central auditory onset responses, and temporal asymmetries in auditory perception. Hearing Research. 167:192–205.

Phillips HN, Blenkmann A, Hughes LE, Bekinschtein TA, Rowe JB. 2015. Hierarchical organization of frontotemporal networks for the prediction of stimuli across multiple dimensions. Journal of Neuroscience. 35:9255–9264.

Phillips HN, Blenkmann A, Hughes LE, Kochen S, Bekinschtein TA, CAN C, Rowe JB. 2016. Convergent evidence for hierarchical prediction networks from human electrocorticography and magnetoencephalography. Cortex. 1–21.

Pickles JO. 2008. An introduction to the physiology of hearing. Bingley.

Picton TW, Hillyard SA, Krausz HI, Galambos R. 1974. Human auditory evoked potentials. I: Evaluation of components. Electroencephalogr Clin Neurophysiol. 36:179–190.

Picton TW, Woods DL, Proulx GB. 1978. Human auditory sustained potentials. II. Stimulus relationships. Electroencephalogr Clin Neurophysiol. 45:198–210.

Pieszek M, Widmann A, Gruber T, Schröger E. 2013. The Human Brain Maintains Contradictory and Redundant Auditory Sensory Predictions. PLoS ONE. 8:e53634–13.

Popham S, Boebinger D, Ellis DPW, Kawahara H, McDermott JH. 2018. Inharmonic speech reveals the role of harmonicity in the cocktail party problem. Nature Communications. 1–13.

Raij T, McEvoy L, Mäkelä JP, Hari R. 1997. Human auditory cortex is activated by omissions of auditory stimuli. Brain Research. 745:134–143.

Rao RP, Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. 2:1–9.

Recasens M, Leung S, Grimm S, Nowak R, Escera C. 2015. Repetition suppression and repetition enhancement underlie auditory memory-trace formation in the human brain: an MEG study. NeuroImage. 108:75–86.

Recasens M, Uhlhaas PJ. 2017. Test-retest reliability of the magnetic mismatch negativity response to sound duration and omission deviants. NeuroImage. 157:184–195.

Riecke L, Sack AT, Schroeder CE. 2015. Endogenous Delta/Theta Sound-Brain Phase Entrainment Accelerates the Buildup of Auditory Streaming. Current Biology. 25:3196–3201.

Rieke, F., Bodnar, D.A. & Bialek, W., 1995. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proceedings of the Royal Society B: Biological Sciences*, 262(1365), pp.259–265.

Rimmele J, Schröger E, Bendixen A. 2012. Age-related changes in the use of regular patterns for auditory scene analysis. Hearing Research. 289:98–107.

Rimmele JM, Morillon B, Poeppel D, Arnal LH. 2018. Proactive Sensing of Periodic and Aperiodic Auditory Patterns. Trends in Cognitive Sciences. 22:870–882.

Roberts TP, Ferrari P, Stufflebeam SM, Poeppel D. 2000. Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. J Clin Neurophysiol. 17:114–129.

Rohenkohl G, Cravo AM, Wyart V, Nobre AC. 2012. Temporal Expectation Improves the Quality of Sensory Information. Journal of Neuroscience. 32:8424–8428.

Rose M, Haider H, Büchel C. 2005. Unconscious detection of implicit expectancies. Journal of Cognitive Neuroscience. 17:918–927.

Rosen S. 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philosophical Transactions of the Royal Society B: Biological Sciences. 336:367–373.

Rubin J, Ulanovsky N, Nelken I, Tishby N. 2016. The Representation of Prediction Error in Auditory Cortex. PLoS Comput Biol. 12:e1005058–28.

Saarinen J, Paavilainen P, Schöger E, Tervaniemi M, Näätänen R. 1992. Representation of abstract attributes of auditory stimuli in the human brain. Neuroreport. 3:1149–1151.

Saffran JR, Aslin RN, Newport EL. 1996. Statistical learning by 8-month-old infants. Science. 274:1926–1928.

Saffran JR, Johnson EK, Aslin RN, Newport EL. 1999. Statistical learning of tone sequences by human infants and adults. Cognition. 70:27–52.

Särelä J, Valpola H. 2005. Denoising Source Separation. Journal of Machine Learning Research. 6:233–272.

Schröger E, Bendixen A, Denham SL, Mill RW, Bőhm TM, Winkler I. 2013. Predictive Regularity Representations in Violation Detection and Auditory Stream Segregation: From Conceptual to Computational Models. Brain Topogr. 27:565–577.

Sedley W, Gander PE, Kumar S, Kovach CK, Oya H, Kawasaki H, Howard MA, Griffiths TD. 2016. Neural signatures of perceptual inference. Elife. 5:797.

Segaert K, Weber K, de Lange FP, Petersson KM, Hagoort P. 2013. The suppression of repetition enhancement: A review of fMRI studies. Neuropsychologia. 51:59–66.

Seger CA, Spiering BJ, Sares AG, Quraini SI, Alpeter C, David J, Thaut MH. 2013. Corticostriatal contributions to musical expectancy perception. Journal of Cognitive Neuroscience. 25:1062–1077.

Sengupta B, Stemmler MB, Friston KJ. 2013. Information and Efficiency in the Nervous System—A Synthesis. PLoS Comput Biol. 9:e1003157–12.

Shamma SA, Micheyl C. 2010. Behind the scenes of auditory perception. Current Opinion in Neurobiology. 20:361–366.

Shannon CE. 1948. A Mathematical Theory of Communication. Bell System Technical Journal. 27:379–423.

Shiner JS, Davison M, Landsberg PT. 1999. Simple measure for complexity. Phys Rev E. 59:1459–1464.

Shipp S. 2016. Neural Elements for Predictive Coding. Front Psychol. 7:611–621.

Simson R, Vaughn HG, Ritter W. 1977. The scalp topography of potentials in auditory and visual discrimination tasks. Electroencephalogr Clin Neurophysiol. 42:528–535.

Singh NC, Theunissen FE. 2003. Modulation spectra of natural sounds and ethological theories of auditory processing. The Journal of the Acoustical Society of America. 114:3394–3411.

Skerritt-Davis B, Elhilali M. 2019. A Model for Statistical Regularity Extraction from Dynamic Sounds. Acta Acustica united with Acustica. 105:1–4.

Slabu L, Escera C, Grimm S, Costa-Faidella J. 2010. Early change detection in humans as revealed by auditory brainstem and middle-latency evoked potentials. Eur J Neurosci. 32:859–865.

Sohoglu E, Chait M. 2016a. Detecting and representing predictable structure during auditory scene analysis. Elife. 5:4182.

Sohoglu E, Chait M. 2016b. Neural dynamics of change detection in crowded acoustic scenes. NeuroImage. 126:164–172.

Solomon SG, Kohn A. 2014. Moving sensory adaptation beyond suppressive effects in single neurons. Curr Biol. 24:R1012–R1022.

Southwell R, Baumann A, Gal C, Barascud N, Friston K, Chait M. 2017. Is predictability salient? A study of attentional capture by auditory patterns. Philos Trans R Soc Lond, B, Biol Sci. 372:20160105–20160111.

Southwell R, Chait M. 2018. Enhanced deviant responses in patterned relative to random sound sequences. Cortex. 109:92–103.

Spratling MW. 2016. A review of predictive coding algorithms. Brain and Cognition.

Stefanics G, Heinzle J, Czigler I, Valentini E, Stephan KE. 2018. Timing of repetition suppression of event-related potentials to unattended objects. Eur J Neurosci. 41:925–10.

Strange BA, Duggins A, Penny W, Dolan RJ, Friston KJ. 2005. Information theory, novelty and hippocampal responses: unpredicted or unpredictable? Neural Networks. 18:225–230.

Summerfield C, de Lange FP. 2014. Expectation in perceptual decision making: neural and computational mechanisms. Nature Publishing Group. 15:745–756.

Summerfield C, Egner T. 2016. Feature-Based Attention and Feature-Based Expectation. Trends in Cognitive Sciences. 1–4.

Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T. 2008. Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci. 11:1004–1006.

Summerfield JJ, Lepsien J, Gitelman DR, Mesulam MM, Nobre AC. 2006. Orienting Attention Based on Long-Term Memory Experience. NEURON. 49:905–916.

Suzuki Y, Takeshima H. 2004. Equal-loudness-level contours for pure tones. The Journal of the Acoustical Society of America. 116:918–933.

Taaseh N, Yaron A, Nelken I. 2011. Stimulus-Specific Adaptation and Deviance Detection in the Rat Auditory Cortex. PLoS ONE. 6:e23369–17.

Tanner D, Morgan-Short K, Luck SJ. 2015. How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. Psychophysiology. 52:997–1009.

Tanner WP Jr, Swets JA. 1954. A decision-making theory of visual detection. Psychol Rev. 61:401–409.

Tecce JJ. 1972. Contingent negative variation (CNV) and psychological processes in man. Psychological Bulletin. 77:73–108.

Teki S, Barascud N, Picard S, Payne C, Griffiths TD, Chait M. 2016. Neural Correlates of Auditory Figure-Ground Segregation Based on Temporal Coherence. Cerebral Cortex. 26:3669–3680.

Teki S, Chait M, Kumar S, Kriegstein von K, Griffiths TD. 2011. Brain Bases for Auditory Stimulus-Driven Figure-Ground Segregation. Journal of Neuroscience. 31:164–171.

Teki S, Chait M, Kumar S, Shamma S, Griffiths TD. 2013. Segregation of complex acoustic scenes based on temporal coherence. Elife. 2013:e00699.

Tervaniemi M, Saarinen J, Paavilainen P, Danilova N, Näätänen R. 1994. Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. Biological Psychology. 38:157–167.

Thomson DJ. 1982. Spectrum Estimation and Harmonic Analysis. Proceedings of the IEEE. 70:1055–1096.

Tillmann B, Poulin-Charronnat B. 2010. Auditory expectations for newly acquired structures. The Quarterly Journal of Experimental Psychology. 63:1646–1664.

Tobia MJ, Iacovella V, Hasson U. 2012. Multiple sensitivity profiles to diversity and transition structure in non-stationary input. NeuroImage. 60:991–1005.

Todorovic A, de Lange FP. 2012. Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. J Neurosci. 32:13389–13395.

Todorovic A, van Ede F, Maris E, de Lange FP. 2011. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. J Neurosci. 31:9118–9123.

Tóth B, Kocsis Z, Háden GP, Szerafin Á, Shinn-Cunningham BG, Winkler I. 2016. EEG signatures accompanying auditory figure-ground segregation. NeuroImage. 141:108–119.

Troebinger L, López JD, Lutti A, Bestmann S, Barnes G. 2014. Discrimination of cortical laminae using MEG. NeuroImage. 102:885–893.

Tukey JW. 1949. Comparing Individual Means in the Analysis of Variance. Biometrics. 5:99.

Turk-Browne NB, Jungé JA, Scholl BJ. 2005. The Automaticity of Visual Statistical Learning. Journal of Experimental Psychology: General. 134:552–564.

Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. 2009. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. Journal of Cognitive Neuroscience. 21:1934–1945.

Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM. 2010. Implicit perceptual anticipation triggered by statistical learning. J Neurosci. 30:11177–11187.

Turner R. 2010. Statistical Models for Natural Sounds.

Ulanovsky N, Las L, Farkas D, Nelken I. 2004. Multiple time scales of adaptation in auditory cortex neurons. J Neurosci. 24:10440–10453.

Ulanovsky N, Las L, Nelken I. 2003. Processing of low-probability sounds by cortical neurons. Nat Neurosci. 6:391–398.

Uluç I, Schmidt TT, Wu Y-H, Blankenburg F. 2018. Content-specific codes of parametric auditory working memory in humans. NeuroImage. 183:254–262.

van Noorden L. 1975. Temporal coherence in the perception of tone sequences.

Vandenberghe M, Schmidt N, Fery P, Cleeremans A. 2006. Can amnesic patients learn without awareness? Neuropsychologia. 44:1629–1641.

Vliegen J, Oxenham AJ. 1999. Sequential stream segregation in the absence of spectral cues. The Journal of the Acoustical Society of America. 105:339–346.

Vossel S, Mathys C, Daunizeau J, Bauer M, Driver J, Friston KJ, Stephan KE. 2014. Spatial Attention, Precision, and Bayesian Inference: A Study of Saccadic Response Speed. Cerebral Cortex. 24:1436–1450.

Wacongne C, Changeux JP, Dehaene S. 2012. A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. Journal of Neuroscience. 32:3665–3678.

Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S. 2011. Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc Natl Acad Sci USA. 108:20754–20759.

Wang D, Clouter A, Chen Q, Shapiro KL, Hanslmayr S. 2018. Single-trial Phase Entrainment of Theta Oscillations in Sensory Regions Predicts Human Associative Memory Performance. bioRxiv. 261487.

Wang L, Uhrig L, Jarraya B, Dehaene S. 2015. Representation of numerical and sequential patterns in macaque and human brains. Curr Biol. 25:1966–1974.

Warren RM, Gardner DA, Brubaker BS, Bashford JA. 1991. Melodic and Nonmelodic Sequences of Tones: Effects of Duration on Perception. Music Perception: An Interdisciplinary Journal. 8:277–289.

Weitzman ED, Kremen H. 1965. Auditory evoked responses during different stages of sleep in man. Electroencephalogr Clin Neurophysiol. 18:65–70.

Widmann A, Schröger E, Maess B. 2015. Digital filter design for electrophysiological data--a practical approach. Journal of Neuroscience Methods. 250:34–46.

Wiley RH, Richards DG. 1978. Physical constraints on acoustic communication in the atmosphere: Implications for the evolution of animal vocalizations. Behav Ecol Sociobiol. 3:69–94.

Winkler I, Denham S, Mill R, Bőhm TM, Bendixen A. 2012. Multistability in auditory stream segregation: a predictive coding view. Philos Trans R Soc Lond, B, Biol Sci. 367:1001–1012.

Winkler I, Denham SL, Nelken I. 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends in Cognitive Sciences. 13:532–540.

Winkler I, Schröger E. 2015. Auditory perceptual objects as generative models: Setting the stage for communication by sound. Brain Lang. 148:1–22.

Winkler I, Takegata R, Sussman E. 2005. Event-related brain potentials reveal multiple stages in the perceptual organization of sound. Cognitive Brain Research. 25:291–299.

Wronka E, Kaiser J, Coenen AML. 2008. The auditory P3 from passive and active three-stimulus oddball paradigm. Acta Neurobiol Exp (Wars). 68:362–372.

Xie Z, Reetzke R, Chandrasekaran B. 2018. Taking Attention Away from the Auditory Modality: Context-dependent Effects on Early Sensory Encoding of Speech. Neuroscience. 384:64–75.

Yabe H, Tervaniemi M, Reinikainen K, Näätänen R. 1997. Temporal window of integration revealed by MMN to sound omission. Neuroreport. 8:1971–1974.

Yao J, Dewald JPA. 2005. Evaluation of different cortical source localization methods using simulated and experimental EEG data. NeuroImage. 25:369–382.

Yarden TS, Nelken I. 2017. Stimulus-specific adaptation in a recurrent network model of primary auditory cortex. PLoS Comput Biol. 13:e1005437–35.

Yaron A, Hershenhoren I, Nelken I. 2012. Sensitivity to Complex Statistical Regularities in Rat Auditory Cortex. NEURON. 76:603–615.

Yu A, Dayan P. 2003. Expected and unexpected uncertainty: ACh and NE in the neocortex. In:. Presented at the Advances in Neural Information Processing Systems.

Yu AJ, Dayan P. 2005. Uncertainty, Neuromodulation, and Attention. Neuron. 46:681–692.

Yvert B, Fischer C, Bertrand O, Pernier J. 2005. Localization of human supratemporal auditory areas from intracerebral auditory evoked potentials using distributed source models. NeuroImage. 28:140–153.

Zatorre RJ, Evans AC, Meyer E. 1994. Neural mechanisms underlying melodic perception and memory for pitch. Journal of Neuroscience. 14:1908–1919.

Zatorre RJ, Salimpoor VN. 2013. From perception to pleasure: Music and its neural substrates. Proc Natl Acad Sci USA. 110:10430–10437.

Zhao J, Al-Aidroos N, Turk-Browne NB. 2013. Attention Is Spontaneously Biased Toward Regularities. Psychological Science. 24:667–677.

Zhao S, Chait M, Dick F, Dayan P, Furukawa S, Liao H-I. 2019. Phasic norepinephrine is a neural interrupt signal for unexpected events in rapidly unfolding sensory sequences - evidence from pupillometry. Biorxiv Preprint. doi:10.1101/466367