FROM SEQUENCES TO COGNITIVE STRUCTURES: NEUROCOMPUTATIONAL MECHANISMS

RYAN MICHAEL CALMUS

Submitted in partial fulfilment of the requirements for
the degree of Doctor of Philosophy

Biosciences Institute
Newcastle University

June 2021

General abstract

Understanding how the brain forms representations of structured information distributed in time is a challenging neuroscientific endeavour, necessitating computationally and neurobiologically informed study. Human neuroimaging evidence demonstrates engagement of a fronto-temporal network, including ventrolateral prefrontal cortex (vlPFC), during language comprehension. Corresponding regions are engaged when processing dependencies between word-like items in Artificial Grammar (AG) paradigms. However, the neurocomputations supporting dependency processing and sequential structure-building are poorly understood. This work aimed to clarify these processes in humans, integrating behavioural, electrophysiological and computational evidence.

I devised a novel auditory AG task to assess simultaneous learning of dependencies between adjacent and non-adjacent items, incorporating learning aids including prosody, feedback, delineated sequence boundaries, staged pre-exposure, and variable intervening items. Behavioural data obtained in 50 healthy adults revealed strongly bimodal performance despite these cues. Notably, however, reaction times revealed sensitivity to the grammar even in low performers. Behavioural and intracranial electrode data was subsequently obtained in 12 neurosurgical patients performing this task. Despite chance behavioural performance, time- and time-frequency domain electrophysiological analysis revealed selective responsiveness to sequence grammaticality in regions including vlPFC. I developed a novel neurocomputational model (VS-BIND: "Vector-symbolic Sequencing of Binding INstantiating Dependencies"), triangulating evidence to clarify putative mechanisms in the fronto-temporal language network. I then undertook multivariate analyses on the AG task neural data, revealing responses compatible with the presence of ordinal codes in vlPFC, consistent with VS-BIND. I also developed a novel method of causal analysis on multivariate patterns, representational Granger causality, capable of detecting flow of distinct representations within the brain. This alluded to top-down transmission of syntactic predictions during the AG task, from vlPFC to auditory cortex, largely in the opposite direction to stimulus encodings, consistent with predictive coding accounts. It finally suggested roles for the temporoparietal junction and frontal operculum during grammaticality processing, congruent with prior literature.

This work provides novel insights into the neurocomputational basis of cognitive structure-building, generating hypotheses for future study, and potentially contributing to AI and translational efforts.

Dedication

This thesis is dedicated to my parents, siblings, friends and late grandparents, who together encouraged in me the strength, confidence and curiosity required to complete it.

It is also dedicated to Ralph Vaughan Williams, Gustav Holst, Sergei Rachmaninoff, Clara Schumann, Ludwig van Beethoven, Martha Argerich, Wolfgang Amadeus Mozart, Edward Elgar, Rachel Portman, James Horner, Sandy Cameron, John Barry, John Williams, Jerry Goldsmith, Fleetwood Mac, Billy Joel, Kate Bush, Rondò Veneziano and every composer, soloist or ensemble whose production of tonal sequences sustained me whilst writing this.

Acknowledgements

The work in Chapters 1 to 6 was supported by: Wellcome Trust PhD Studentship (RC), Wellcome Trust Investigator Award (WT092606AIA), BBSRC (BB/J009849/1), European Research Council (ERC CoG, MECHIDENT) and NIH (R01-DC04290).

Chapter 2: I would like to thank Oana Morteanu for her hard work and immense efficiency in behavioural testing. I would also like to thank my academic supervisors for their invaluable assistance in discussing the parameters of the paradigm.

Chapter 3: I would like to thank the wonderful team at Iowa's Human Brain Research Laboratory, including: Zsuzsanna Kocsis, Beau Snoad, Ariane Rhone, Haiming Chen and Phillip Gander for illuminating discussions, collecting data, or assisting me during data collection; Christopher Kovach and Kirill Nourski for helpful advice; Matt Howard and Hiroto Kawasaki for supporting such a vibrant collaboration; and the whole team once again for their friendship, feedback and kindness. I would also like to thank all of the other friends and colleagues I made in Iowa, including Alex Billig, Araceli Ramírez-Cárdenas and McCall Sarrett, who, like the rest of the lab, shared freely of the fountains of their knowledge and made me feel truly welcome.

Chapter 4: I would like to thank my co-authors, and Elizabeth Buffalo, Pascal Fries, Karl Friston, Tim Griffiths, David Poeppel, Mark Stokes and Chris Summerfield for inspiring discussions on complex combinatorial binding.

I would also once again like to thank my supervisors, Christopher I. Petkov, Yukiko Kikuchi and Benjamin Wilson, for their boundless optimism, advice, feedback and friendship, as well as my desk- and office-mates throughout the ages, including Cassandra Smith, David Hunter, Jennifer Nacef, Heather Slater, Gillian Pepper, Alice Milne, Joe Necus, Ross Muers, Holly Jenkins and Cody McCants. I could not have hoped for better mentors, colleagues, and friends, and their support means more than I can express with mere words.

Finally, I would like to emphatically thank my parents for still expressing pride in my achievements despite my seemingly perpetual studenthood, and for their patience in listening to me witter on endlessly about the brain and everything else.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AGL	Artificial grammar learning
BA	Brodmann area
CI	Confidence interval
DLPFC	Dorsolateral prefrontal cortex
DRD	Dynamic representational dissimilarity
ECoG	Electrocorticography
EEG	Electroencephalography
fMRI	Functional magnetic resonance imaging
FOP	Frontal operculum
HG	Heschl's gyrus
IFG	Inferior frontal gyrus
MNI	Montreal Neurological Institute
MRI	Magnetic resonance imaging
pMDS	Procrustes multidimensional scaling
RDM	Representational dissimilarity matrix
ROI	Region of interest
RSA	Representational similarity analysis
STG	Superior temporal gyrus
TP	Transition probability
TPJ	Temporoparietal junction
VLPFC	Ventrolateral prefrontal cortex
VSA	Vector symbolic architecture
VS-BIND	Vector-symbolic Sequencing of Binding INstantiating Dependencies

List of Symbols

$\otimes$	Circular convolution
$\neg$	Involution
$\mathbb{R}$	The set of all real numbers
$\mathbb{R}^n$	The set of all n-dimensional real-valued vectors

Chapter 1. Introduction

Whatever Nature undertakes, she can only accomplish it in a sequence.

(Johann Wolfgang von Goethe, wilfully misused.)

1.1 The universality of temporal structure

Whilst Johann von Goethe was alluding to the stepwise character of evolution by natural selection, there is a broader truth to his words: sequences are everywhere. Earth is filled with predictable sequential patterns, from the rising and setting of the sun each day, to the movement of the tides and the lifecycles of its lively inhabitants. These sequential patterns can be useful. In particular, a species with the luxury of a brain may attempt to minimise risk, and maximise reward, by exploiting the sequential statistical structure in its environment, making predictions of the future that can modulate arousal and attention (Hansen & Pearce, 2014).

Humans in particular are adept at manipulating and ancitipating sequence structure, as exemplified by our unique ability to communicate using speech and language. We produce and comprehend written and spoken streams of words, each of which has meaning (*semantics*), and which are organised according to particular combinatorial rules (*syntax*) that support the transmission of a theoretical infinitude of meaningful sentences (Harris, 1957; Chomsky, 1957). Just as an ability to predict the future by learning simple sequencing relationships provides inherent rewards (Hansen & Pearce, 2014), it has been theorised by some that linguistic communication is an inherently adaptive trait, which may be specialised in humans as a result of evolution by natural selection and potentially also as one of the key advantages provided by our relatively large brains (Schoenemann, 2009). However, though some of the systems supporting language may be completely human-unique specialisations (Hauser et al., 2002), there is evidence that many of its neurobiological substrates may constitute advancements on evolutionarily conserved functionality, forming part of an ancestral system of domain-general relevance to sequence processing (Kikuchi et al., 2018; Wilson, Kikuchi, et al., 2015; Petkov & Wilson, 2012; Hauser et al., 2002). Certainly, our ability to communicate using language is but one example of a number of remarkable feats we are able to accomplish that seemingly all relate to a general capacity to manipulate sequences. This

capacity has been considered relevant to the production and consumption of music (Fitch & Martins, 2014; Jeon, 2014; Koelsch et al., 2013; Koelsch & Siebel, 2005); to the undertaking of arithmetic operations (Dehaene et al., 2015; Friedrich & Friederici, 2009; Kong et al., 2005; Dehaene et al., 2004); and, more generally, to the production of plans and complex action sequences (Rouault & Koechlin, 2018; Fitch & Martins, 2014; Bickerton & Szathmáry, 2009; Dehaene & Changeux, 1997), as well as to cognition in general (Bickerton & Szathmáry, 2009; Jeon, 2014).

What does it mean to process a sequence, or to learn sequence structure? Here, we can appreciate the commonalities between the domains above: critically, even though words, musical notes, arithmetic operations and actions may be presented to us as linear sequences that unfold over time, we can learn to appreciate in them an innate structure that can differ from a simple linear ordering of all the presented items or events. This includes relationships between specific neighbouring items (*adjacent dependencies*); between items separated over time (*non-adjacent dependencies*); and relationships between perceived superordinate and subordinate parts (*hierarchical dependencies*). Such structure might in actuality only be present within the sequence at the most abstract conceptual level, but it can nevertheless appear obvious to us. In language syntax, complex structure, incorporating adjacent, non-adjacent and hierarchical dependencies, is ubiquitous (Deocampo et al., 2019), present even within this single sentence in the form of nested phrases. However, not all dependencies are created equal; unifying different kinds of dependencies requires different levels of computational ability. By way of an example, we may conceive of a sequence as a set of states, between which we may transition with a certain frequency (see Figure 1.1). Tabulating these states and the observed frequency of specific transitions produces a graphical understanding of the likelihood of transitioning between each pair of items. This is known as a *Markov model* (Rabiner & Juang, 1986). Here, we depict two Markov models as graphs, each "trained" on a set of sequences conforming to a set of specific dependencies, where important items (**A1**, **A2**, **B1**, **B2**) may be next to each other (**xA1B1** and **xA2B2** permitted, which are adjacent dependencies) or separated by an uninformative intervening item (**X**; valid sequences **A1xB1** and **A2xB2**; non-adjacent dependencies). A Markov model can take account of a history of multiple items when tabulating state transitions. Within Figure 1.1, it can be seen that a simple *first-order* model, which only takes account of a history of one item (left), permits invalid sequence structures to be generated (observed by tracing all routes along the graph). By contrast, a *second-order* model (right)

maps all valid sequences and permits no invalid sequences. In this way, we can see that different dependency types impose different computational burdens on the brain. By studying the limits of dependency processing abilities within and beyond language, we may therefore improve our characterisation of the mechanistic constraints of the human structured sequence processing system, which may in turn constrain hypotheses on human-unique cognitive abilities (Fitch & Martins, 2014; Rosenbaum et al., 2007; Hauser et al., 2002; Lashley, 1951).

As we have seen, the human aptitude for learning and manipulating abstract structure is exemplified by our facility with language, but seemingly also of relevance to our performance in an array of cognitive domains. In particular, three strongly interrelated abilities may subserve performance across these domains. Firstly, we appear to have a particular strength for establishing robust encodings of stimulus ordering and category. Relatively recently, it has been argued that the robustness of these encodings is potentially human-unique, and could be one of several defining features supporting human sequence processing abilities (Ghirlanda et al., 2017). Secondly, we are adept at learning covert structure between items on the basis of the statistical relations between them (known as *statistical learning*), especially items unfolding over time (*sequence learning*, or *structured sequence learning*). We can learn this structure from mere exposure (known as *incidental learning;* Saffran et al., 1997), seemingly without awareness of what we have learned (*implicit learning*; Reber, 1967). In particular, we can learn hierarchical sequencing relationships, an ability which has yet to be clearly established in other species (ten Cate et al., 2020). Thirdly, during and after learning, and presumably supported by robust item encodings, we appear to be able to efficiently transform between different mental structures. This includes transforming between linear, sequential representations of the world, such as words or actions that unfold serially over time, and complex cognitive representations of abstract structure. It also includes combinatorial operations that we may use to synthesise an infinitude of new structures. Specifically, this has been posited to involve the encoding of abstract mental "trees" of information (Dehaene et al., 2015; Chomsky, 1956). The degree to which combination of constituents into trees represents the source of our linguistic abilities is debated (Pinker & Jackendoff, 2005; Chomsky, 1995), but a tree-like conceptual structure has been mooted as an essential component in theories of language and cognition (Dehaene et al., 2015).

The study of these three related abilities spans a diverse array of distinct research domains, some of which are only recently beginning to be directly reconciled (Christiansen, 2019). As we shall

see, sequence learning tasks have informed understanding of human abilities to discern covert temporal structure and establish robust serial orderings across domains, and have also supported the identification of possible shared neural substrates (Friederici, 2011). However, despite advances in the field, we still lack a concrete understanding of the neurocomputational bases of sequence processing, a clear understanding of those mechanisms shared between language and non-linguistic sequence processing, or indeed a clear picture of those factors that may help or hinder sequence learning performance (ten Cate et al., 2020). In this chapter I will discuss how experimental sequence learning tasks have provided a window on critical mechanisms relevant to the appreciation of sequence structure, and to higher cognition. I will also present existing evidence that structured sequence processing evokes a plurality of neural encodings representing orthogonal features. Subsequently, I will consider the neurocomputational implications of these and other prior findings, reflecting on how we may constrain hypotheses on the mechanisms subserving language and cognition in general. Finally, I will summarise the limits of our understanding of human sequence processing abilities and representations, and motivate key research goals that must be achieved in order to provide clarification of these aspects.

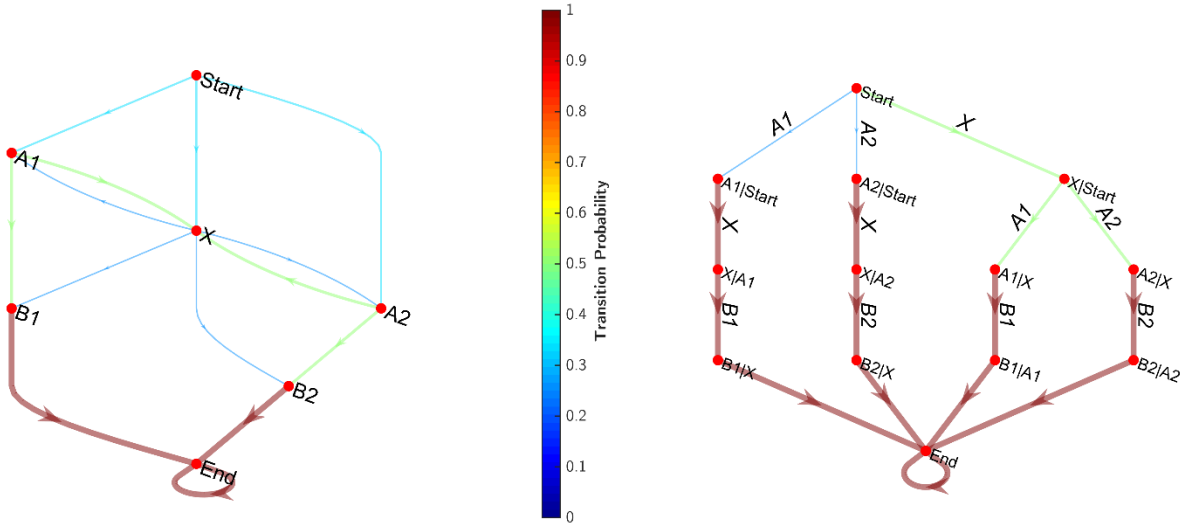


Figure 1.1: **State-transition graphs of first-order (left) and second-order (right) Markov models produced by analysing the frequency of transitions in a mixed dependency grammar.** The graphs depict legal transitions between items in a sequence unfolding over time, as detected by two types of Markov model. Nodes represent states, which, here, relate to the most recently presented item(s) in the sequence. Edges depict non-zero transition probabilities (TPs). Wider edges and warmer colours indicate higher TPs. On the left, each node is labelled according to the single most recent item. On the right, where the states of the model are more complex, the newly presented item is shown as an edge label. The grammar consists of relationships between two important items, **A** and **B**, which can be adjacent (**xAB**) or nonadjacent (**AxB**). **X** is uninformative, but **A** and **B** must match (using our notation, $\mathbf{A_1}$ must eventually be followed by $\mathbf{B_1}$, and $\mathbf{A_2}$ by $\mathbf{B_2}$). The first-order Markov model can only incorporate a history of a single item, whereas the second-order model takes account of a history of two items. As a result, the first-order model is unable, without the support of additional mechanisms, to encapsulate the fact that $\mathbf{A_1xB_1}$ and $\mathbf{A_2xB_2}$ are valid sequences, whilst $\mathbf{A_1xB_2}$ and $\mathbf{A_2xB_1}$ are not. The model also supports many other illegal sequences, such as $\mathbf{A_1xA_2xA_1xA_1}$, for example. However, it correctly rejects all incorrect **xAB** sequences. By contrast, the second-order model correctly allows all valid transitions and rejects all invalid ones, for both adjacent and non-adjacent sequences.

1.2 The experimental study of sequence learning

A great variety of paradigms have indirectly informed our understanding of how sequences and sequence items are encoded, including *recall tasks* (Klein et al., 2005), which assess abilities to store and recall sets of ordered or unordered items; *cross-situational learning tasks* (Kachergis et al., 2014), which assess abilities to learn the meaning of novel "words" by establishing word-referent

pairings over time; *categorisation tasks* (Alfonso-Reese et al., 2002), which assess abilities to categorise stimuli, and may inform our understanding of categorical sequence encodings (Dehaene et al., 2015); and *oddball tasks* (Herrmann & Knight, 2001), in which the presentation of repetitive stimuli is occasionally interrupted by a deviant stimulus. These more general cognitive tasks can help us to constrain hypotheses on sequence encoding. However, a subset of paradigms are able to provide more direct evidence of a subject's ability to acquire structural relationships between elements of serially ordered stimuli (that is, they at least partially involve *grammar acquisition*).

Below, I restrict further discussion to paradigms in which grammar acquisition is a prominent aim or likely outcome of the task: *serial reaction time tasks*, *artificial grammar learning tasks*, *spatial navigation tasks, natural language* and *miniature language tasks*. These tasks vary in their complexity, with some assessing learning of complex behaviours using ecologically valid stimuli, and others illuminating our understanding of sequence processing through a narrower and even deliberately artificial focus. However, all are united in their potential to inform understanding of generic principles and mechanisms supporting the acquisition and encoding of structural relationships in sequences.

1.2.1 Serial reaction time tasks

The serial reaction time (SRT) task (Nissen & Bullemer, 1987) measures a participant's ability to reproduce a specific presented sequence. In the SRT task, a sequence of stably ordered cues is initially presented to the participant (typically visually, on a computer screen, at one of four distinct locations), each requiring a response. Generally, the response takes the form of a button press, where there are as many buttons as cue locations and the participant must respond as quickly and as accurately as possible to the onscreen stimulus by pressing the correspondingly positioned button. When the corresponding button is pressed, the sequence continues, requiring another response; and so on. Over time, as the participant acquires knowledge of the sequence structure, their reaction times will decrease, as they anticipate the upcoming cue position and respond accordingly. The cue ordering is subsequently randomised, whilst response times continue to be measured. Analysis of overall response times, and comparison of response times between the two key phases of the task (ordered and random cues), can yield valuable insights not only into the participant's ability to learn the visuomotor association, but also to learn, and ultimately predict, the structure of the presented stable sequences (Robertson, 2007). Along with artificial grammar learning tasks (see below), SRT

tasks have been suggested as a means of measuring human implicit learning capabilities (Esser & Haider, 2017) as well as providing insights into sequence learning abilities (Sense & Rijn, 2018; Schwarb & Schumacher, 2012; Song et al., 2008). Moreover, in addition to the classical visual task, auditory SRT tasks have also been devised (Terry et al., 2016; Zhuang et al., 1998).

As in the case of spatial navigation (see 1.2.2) and artificial grammar learning tasks (see 1.2.4), but unlike natural language tasks (see 1.2.3), SRT tasks have the advantage that they are suitable for deployment in non-verbal or even infant humans (Koch et al., 2020) as well as non-human species, for example rhesus macaques (Procyk et al., 2000; Heimbauer et al., 2012). This makes them one of a number of tasks that permit the comparative study of sequence learning abilities. More specifically, it allows judicious deployment in certain animal models where it may be possible to use invasive methods of neural recording unsuitable for use in humans. This permits the researcher to obtain simultaneous recordings that may better elucidate specific neural mechanisms supporting sequence processing.

The SRT task is clearly an elegant and versatile paradigm. However, as few naturalistic scenarios impose similar task demands, it can be argued that they lack ecological validity as a means of studying sequence learning beyond the laboratory. For example, because the SRT task requires subjects to rapidly interleave their responses with sequence item presentation, even auditory versions arguably provide an environment somewhat unlike that encountered during the early development of language (Saffran et al., 1997). This particular limitation can be overcome by using an artificial grammar learning task (see 1.2.4). However, alternatively, to maximise ecological validity, one may look to tasks involving fewer constraints or invoking more complex behaviours, such as natural language or spatial navigation tasks.

1.2.2 Spatial navigation tasks

Spatial navigation tasks encompass a wide variety of paradigms in which a subject, human or non-human, must navigate a spatial environment to reach some defined goal, or undertake self-guided exploration of a space. Such tasks can require the individual to navigate through naturalistic or artificial environments, which the subject might experience either in the real-world or, increasingly, through virtual reality. Subjects complete such tasks by engaging a variety of dynamic, multisensory cognitive processes, for example by dynamically updating an estimate of their relative

position in the environment (*egocentric position*) using sensory cues generated by the subject's own movements (*path integration*); by combining distance and directional estimates to render a cognitive map of the positions of landmarks in the environment (*allocentric position*); by using stored knowledge of the relative positions of landmarks to orient themselves or make judgements of relative direction between them; and by undertaking abstract planning and decision-making on the optimal path to take to reach a goal (see Ekstrom et al., 2018).

Although the demands of spatial navigation tasks may not at first appear to relate to temporal sequence structure, it has been proposed that they nevertheless engage key processes relevant to overt sequence learning tasks, to the extent that it has even been suggested that navigation and sequence parsing are instances of the same generic computation (Bartlett & Kazakov, 2005). For example, during spatial navigation, as described above, an individual may instantiate a spatial map of specific locations, and navigate with agency between these locations by using knowledge of appropriate transitions between them, potentially also encoding and retrieving serially ordered lists of those locations that form a path to a goal or subgoal. In a similar way, in order to undertake a sequence learning task, one might solve the problem by instantiating an abstract map of specific sequence items, and "navigate" between these items by using knowledge of appropriate probabilistic transitions, likewise storing serial orderings of items or "chunks" of items as necessary. Taking such a view, some have theorised that all complex behaviours, including language, strongly revolve around the problem of dynamic navigation through physical or abstract space (Edelman, 2017).

Empirical evidence also supports the notion that shared neuronal substrates underpin both sequence learning and spatial navigation, chief among these being the hippocampus (see 1.3), long known to have key roles in both a spatial memory system (O'Keefe, 1976) and the encoding and retrieval of episodic memory (Eichenbaum, 2013). The hippocampus is strongly implicated in domain-general sequencing computations (Buzsáki & Tingley, 2018; Schuck & Niv, 2019), as well as the mapping of both physical and abstract space (Tavares et al., 2015; Park et al., 2020; Nieh et al., 2021).

A great variety of spatial navigation tasks have been undertaken in order to illuminate the functioning of specific components of the spatial memory system (see Ekstrom et al., 2018), but in general, spatial navigation tasks have certain advantages and disadvantages as a means of studying

sequence learning. More overt sequence learning tasks, such as SRT tasks (see 1.2.1) and artificial grammar learning tasks (see 1.2.4), may have an advantage over spatial navigation tasks in restricting stimuli to temporal sequences, typically presented through a specific sensory modality, which simplifies analysis and the testing of formal computational hypotheses. Conversely, however, it may be argued that spatial navigation tasks have more ecological validity than these non-linguistic sequencing tasks, by engaging a host of complex cognitive processes in naturalistic environments. Additionally, when compared to natural language tasks, spatial navigation tasks have the advantage that they are readily deployed in non-human species. This allows for comparative study, as well as the simultaneous collection of rich neural and behavioural data such as that collected by O'Keefe (1976), which famously established the existence of hippocampal *place cells* representing the position of freely moving rats in a maze. However, when studying human sequence processing specifically, it is difficult to ignore what appears to be the epitome of human-unique temporal sequence learning: natural language.

1.2.3 Natural language and miniature language tasks

The study of natural language encompasses an enormous breadth of research within and beyond neuroscience that has the potential to shed light on many essential principles of human cognition beyond the scope of this discussion. However, as a form of communication rooted in sequential structure, language more specifically offers an opportunity to study generic principles of human sequence processing using naturalistic, ecologically valid stimuli.

In theory, there are a number of advantages to using natural language to explore mechanisms of sequence processing. Natural language is highly salient to most human subjects; language skills represent the culmination of years of development and learning, such that we tend to display an innate awareness of synactic correctness and extreme sensitivity to manipulations of sentence structure; and natural language includes meaning, or semantics, and complex morphology that allows us to study syntax as it interacts with other features. This latter point may be especially important, as we do not learn natural language syntax in isolation from semantics, and neither aspect is fully subordinate to the other in development, with some combinatorial semantic processes seemingly available before the full developmental maturation of syntactic processes (Morgan et al., 2020). Whilst the relative influences of semantics and syntax on comprehension vary throughout our lifetimes (Wu et al., 2016), and under different task demands, the two are demonstrably intertwined

when determining the statistical likelihood of specific sequences. For example, the learning of previously unseen miniature language syntax (see below) is aided by the inclusion of semantic content (Poletiek et al., 2021). Unsurprisingly, therefore, natural language tasks have been used extensively to study the neurobiological bases of semantic processing, with many tasks seeking to tease apart the interactions of meaning and structure, and their relative contributions to observed neural responses (Friederici, 2011; Price et al., 2016; Angrilli et al., 2002; Münte et al., 1993; Kutas & Hillyard, 1983).

However, some of the advantages to natural language tasks listed above are also potential disadvantages. Firstly, because a long period of development and learning underpins our linguistic abilities, it is not feasible to observe the full trajectory of natural language learning within a small number of experimental sessions. To overcome this, miniature language learning studies (see below) and non-linguistic sequencing tasks (see 1.2.1 and 1.2.4) theoretically afford the experimenter a view of sequence learning from first principles in the relative absence of prior knowledge (although even under these tasks the subject cannot be regarded as being a completely "blank slate"; see 1.2.4). Secondly, whilst knowledge of language semantics can aid learning of syntax and vice versa, the presence of meaning in everyday utterances introduces an unavoidable confound into any task specifically exploring responses to sequence structure manipulations. This is likely to be problematic for the neuroscientist, because semantic processing engages a host of brain regions, some of which may vary considerably with the meaning of the presented words (Hertrich et al., 2020), and others of which may also overlap spatially with syntax-sensitive areas (Petersson et al., 2012; Petkov & Wilson, 2012; Fuji et al., 2016).

A close alternative to natural language – and one that potentially allows a greater degree of control over the subject's response – is the *miniature language* or *artificial language* task. In these tasks, subjects are exposed to specially designed novel "languages" that have known syntax and in which the vocabulary has meaning (for example, Friederici et al., 2002; Cross et al., 2020; Poletiek et al., 2021). Unlike tasks using stimuli in the subject's first language, however, and more like artificial grammar (see below) or serial reaction time tasks (see 1.2.1), miniature language tasks allow the researcher to study the learning of meaning and structure from a relatively clear baseline, and/or with intentional restrictions. Friederici et al. (2002), for example, used such a paradigm in adults to show that an artificial language, *BROCANTO*, could elicit extremely similar syntactic responses to those

provoked by the subjects' native language, though a body of second-language-learning tasks had previously suggested that strongly native-like-responses could only be elicited by languages learned in a critical period of childhood development (Weber-Fox & Neville, 1996). Friederici et al. theorised that this was achievable because the restricted size of the artificial language both recapitulated childhood restrictions on information processing capacity and allowed adults to achieve a high level of proficiency on the task (Friederici et al., 2002).

Natural and miniature language tasks are commonly regarded as boasting substantial ecological validity compared to non-linguistic sequence learning tasks (Cross et al., 2020). Like spatial navigation tasks, they involve complex behaviours and salient, naturalistic stimuli and task demands familiar to us in everyday life. However, it may not necessarily be sensible to assume that this is the case in neuroimaging studies. This is because the majority of existing neuroscientific natural language studies typically present deliberately over-simplified instances of day-to-day constructs, measuring the processing of specific classes of words or sentences, in order to avoid confounds and ensure the feasibility of analysis (Hauk & Weiss, 2020). However, this landscape is gradually changing as researchers increasingly promote analytical techniques and datasets focussed on freeform natural language stimuli (Hamilton & Huth, 2020; Himmelstoss et al., 2020; Kaestner et al., 2020).

As previously described, natural and miniature language tasks are not the only ecologically relevant way to study structured sequence learning. Multiple alternative paradigms have been devised to help us better understand how we learn sequential representations under controlled conditions, whilst eliminating the effects of semantic confounds (see 1.2.1 and Miller, 1958; Reber, 1967; Berry & Broadbent, 1984; Nissen & Bullemer, 1987). Lacking meaning, as they do, such paradigms have the advantage over natural language tasks that they can also be conducted in non-verbal humans, or even non-human species, supporting research that can inform our understanding of the ontogeny or phylogeny of specific neural mechanisms. In particular, the use of non-linguistic tasks in humans allow us to bridge a gap between humans and animal models, enabling comparisons to be drawn between species to inform understanding of potentially shared cognitive capabilities that pre-date the human faculty for language (Petkov & ten Cate, 2020). Arguably the most popular of these non-linguistic sequence learning paradigms is the artificial grammar learning task, which we discuss in the following section.

1.2.4 Artificial grammar learning tasks

The artificial grammar learning (AGL) paradigm (Miller, 1958; Reber, 1967) tests subjects on their ability to learn structured sequences that conform to a set of covert ordering rules, known as an *artificial grammar*. This involves initially habituating participants only to sequences that conform to this artificial grammar, during an *exposure phase*. The grammar may be any set of ordering rules (for example, "**Q** must always follow **P**"), which are then used to generate the set of exposure sequences. Subsequently, during a *testing phase*, participants are assessed on their ability to distinguish novel "grammatical" sequences (which also conform to the ordering rules) from "ungrammatical" sequences (which do not). Beyond this, there are many variations on the AGL paradigm, particularly in terms of the modality of the presented stimuli and the response measures used in the testing phase. For example, in terms of modality, the classical AGL task, devised by Reber (1967), presented written verbal sequences, but AGL tasks have also been performed using sequences of spoken nonsense words (Saffran, Aslin, et al., 1996), tone sequences (Saffran et al., 1999), visual shape sequences (Stobbe et al., 2012; Fiser & Aslin, 2002), and even observed dance routines (Opacic et al., 2009). Similarly, response measures used to determine learning of the grammar include, for example, explicit two-alternative forced-choice responses assessing perceived grammaticality or familiarity; implicit reaction time measures, head-turn preferences and other orienting responses; neuroimaging effects, or combinations of these (Batterink & Paller, 2017; Attaheri et al., 2015; Wilson et al., 2013; Saffran, Aslin, et al., 1996; Reber, 1967).

Minimal instruction need be provided under an AGL task. This fact, in conjunction with the flexibility of the paradigm, means that it has been possible to use AGL tasks to reveal sequence learning effects in diverse populations, including healthy or language-impaired human adults (Cope et al., 2017; Schiff et al., 2017; Petersson et al., 2012; Gómez, 2002; Saffran et al., 1999) and children (Witt & Vinter, 2012; Evans et al., 2009), pre-linguistic infants (Gómez, 2002; Saffran et al., 1999; Saffran, Aslin, et al., 1996), and even in a host of other species, including, for example, chimpanzees (Sonnweber et al., 2015), macaques, marmosets (Wilson et al., 2013), rats (Toro & Trobalón, 2005), budgerigars, zebra finches (Spierings & ten Cate, 2016) and pigeons (Herbranson & Shimp, 2008). Crucially, as suggested by the diversity of this list, AGL tasks can be performed in any population that lacks the capacity for language, in part because the stimuli are not semantically meaningful and therefore no comprehension is required. In many respects, serial reaction time (SRT) tasks offer a

similar advantage (see 1.2.1), although it is likely easier to adapt the AGL than the SRT task to completely passive use in non-language-capable populations (for example by recording eye-tracking data), since fewer responses are required from the subject under the AGL paradigm.

As discussed in section 1.2.3, the semantics of natural language can be viewed as a confounding variable. Although AGL tasks can be used to study structural relationships with similarities to language syntax (Petkov & Wilson, 2012) in naturalistic speech stimuli (Milne et al., 2018; Wilson et al., 2013; Gómez, 2002; Saffran, Aslin, et al., 1996), they allow the experimenter to do so in relative isolation from other components of language, including semantics (Petersson et al., 2012). As previously discussed, this can be especially desirable when studying the neural responses of human subjects, because language invokes many stimulus-dependent forms of processing that may contaminate a view of any structure-specific effects (Hertrich et al., 2020). It also potentially means that findings under AGL tasks can be more readily generalised to explain a broad range of non-linguistic processes than can linguistic tasks. However, AGL tasks are frequently suggested to provide a window on language. From this perspective, a lack of semantics is disadvantageous since it means that learning cannot occur through the same complex syntactic-semantic interplay that natural language acquisition likely involves (Morgan et al., 2020; Wu et al., 2016). Furthermore, although performance on natural language tasks has been shown to correlate with performance on AGL tasks (see for example Frost et al., 2015), it may not be appropriate to conflate natural language syntax and artificial grammars, since the former involves highly established learning of patterns abstracted across categories representing thousands of tokens (words), whereas AGL tasks involve rapidly learned associations between small numbers of distinct tokens.

There are other potential disadvantages to artificial grammar learning tasks, relative to alternative approaches. Firstly, although human AGL subjects are typically presented with minimal instructions, the typical AGL task is sufficiently unguided that the wording of these instructions may affect what the subject perceives to be the demands of the task. Previous studies have suggested that AGL task results can be surprisingly robust to instruction effects in healthy adults (Forkstam et al., 2008) although, conversely, adults with developmental dyslexia have been shown to perform worse than healthy controls when provided with scant, implicit instructions, and only match control performance when provided with explicit task instructions (Kahta & Schiff, 2016). Depending on the population, therefore, AGL results may be quite vulnerable to instruction effects. Secondly,

AGL stimuli arguably lack the salience of natural language stimuli or novel spatial environments, making the paradigm fatiguing and dull for subjects. This can be mitigated, for example through "gamification" of the task (Hall et al., 2018), although in such tasks the presentation of additional entertaining media has the potential to contaminate neural responses. Thirdly, although AGL tasks, like miniature language tasks, are commonly used as a means of studying sequence learning in the absence of extensive prior knowledge, the subjects performing these tasks cannot be regarded as starting from a completely blank slate, since first language knowledge is known to bias the learning of artificial languages (Onnis & Thiessen, 2013; LaCross, 2015). To some extent, the experimenter can mitigate biases resulting from the poor choice of non-linguistic stimuli, for example nonsense units that resemble meaningful words – but other biases, such as a first-language bias towards learning a particular sequence construction, are less easily controlled. These factors must surely affect other proto- and non-linguistic tasks, but our knowledge of such biases suggests the imposition of a limit on how far we generalise from AGL task results in a specific sample to the global population. There are thus a number of limitations to be borne in mind when using AGL tasks. As discussed, however, despite their potential pitfalls, AGL paradigms also have particular strengths. For example, unlike the majority of spatial navigation tasks, they draw upon a defined grammar and permit the investigation of temporal sequence processing in the absence of spatial cues. Like serial reaction time tasks (see 1.2.1), they permit the experimenter to investigate principles relevant to language in isolation from the influence of semantics, and in a wide range of populations (Udden & Männel, 2018). As we have seen, though this may harm their ecological validity as completely faithful reproductions of natural language development, it arguably makes them more useful as neuroimaging tasks.

Despite extensive use of the AGL paradigm in humans, we have yet to fully characterise the limits of human dependency learning abilities under these tasks (ten Cate et al., 2020; Gebhart et al., 2009). In particular, despite the prevalence of non-adjacent dependencies in natural language, AGL studies of non-adjacent dependency learning have yielded inconsistent results in adults. It has been suggested that this inconsistency arises because the ability to learn non-adjacent dependencies is more selective compared to adjacent dependency learning (Gebhart et al., 2009), but this requires clarification. In particular, by establishing fundamental limits on dependency learning, we might impose useful constraints on neurocomputational accounts of sequence processing.

1.3 Anatomical substrates of sequence processing

Neuroimaging studies have revealed that a core system, the *frontotemporal language network* (Friederici, 2011), responds similarly to manipulations of complexity of both language syntax and non-linguistic sequence structure. This frontotemporal network (or *core language network*; Hertrich et al., 2020) comprises areas in ventrolateral prefrontal, temporal and temporoparietal cortices, including, in particular, subregions of inferior frontal gyrus ("Broca's area"; Broca, 1861) and "Wernicke's area" (Wernicke, 1874; Friederici, 2011). These two eponymous regions have especially historic associations with language. Broca's area, for example, has been a source of fascination and intense study for 160 years, one of the first modern attempts to localise specific cerebral functions in the human brain. Louis Victor Leborgne or, as he is most famously known, "Tan", was a millinery mould-maker and patient of Pierre Paul Broca who, at the age of 30, lost the ability to speak, save for the one word that gave him his nickname: "tan" (Mohammed et al., 2018). By localising to left inferior frontal gyrus the lesion that appeared to give Tan his name, Broca proposed for it a critical role in language – initiating a cascade of research that remains strongly relevant to this day, and to which we shall return shortly.

As well as cortical regions, subcortical areas have been implicated in the processing of sequences. Perhaps most prominent among these is the hippocampus, which is known to have a critical role in both spatial (for example, O'Keefe, 1976; see also 1.2.2) and episodic memory formation (for example, Eichenbaum, 2013). More specifically, recent studies provide clear evidence of an online role for the hippocampus in sequence learning tasks (Opitz & Friederici, 2003; Schapiro et al., 2014; Covington et al., 2018; Jablonowski et al., 2018; Kepinska et al., 2018; Henin et al., 2021) as well as in the acquisition and comprehension of semantically meaningful language (Breitenstein et al., 2005; Duff & Brown-Schmidt, 2012; Piai et al., 2016).

Despite strong evidence of its engagement, the exact role of the hippocampus in sequence learning is presently unclear, although it has been demonstrated to contribute to both working memory for sequential content (Faraco et al., 2011; Kumar et al., 2016; Roberts et al., 2018; Shahbaba et al., 2019) as well as to the implicit learning process itself (Schapiro et al., 2014; Covington et al., 2018). Moreover, studies have shown that the hippocampus engages in complex interactions with inferior frontal gyrus during implicit sequence learning tasks (Opitz & Friederici,

2003, 2004; Gheysen et al., 2010; Kepinska et al., 2018). Of note, a study by Opitz & Friederici (2003), in which subjects were required to learn the structure of the *BROCANTO* language (see 1.2.3), suggested that the burden of processing initially fell more heavily on the hippocampus, but transitioned towards increased inferior frontal gyrus activation as the subject's proficiency increased. However, perhaps most importantly, a great deal of research has suggested that the computations it performs generalise across the spatial and non-spatial domains (for example, Howard & Eichenbaum, 2013; Buzsáki & Tingley, 2018; Theves et al., 2019; Whittington et al., 2020; Nieh et al., 2021), instantiating codes that incorporate organisational features in both time and space (as demonstrated in, for example, Kraus et al., 2013). On this basis, it can be seen that forming a coherent domain-general account of hippocampal processing has the potential to constrain hypotheses on hippocampal function, including hypotheses on the hippocampal role in sequence processing (see 1.4).

Although the role of the hippocampus in sequence processing remains imperfectly understood, the neuro-computational specifics of its functioning are still arguably better understood than those of the inferior frontal gyrus. Despite its longstanding general associations with language, highly specific roles for the inferior frontal gyrus have yet to be determined, and the proposed possibilities are many and various. These include the resolution of perceptual conflict (Hsu et al., 2017); action observation and execution (Molenberghs et al., 2012; Pulvermüller & Fadiga, 2010; Rizzolatti & Arbib, 1998); working memory (Yan et al., 2021; Rogalsky et al., 2008; Smith & Jonides, 1999; Caplan & Waters, 1999); and processes of syntactic unification or structure-building (Matchin & Hickok, 2020; Hagoort, 2005; Friederici, 2002; Grodzinsky, 2000). Notably, however, engagement of left ventrolateral prefrontal cortical (vlPFC) regions appears to vary with the complexity of natural language syntax or sequence structure (Friederici, Bahlmann, et al., 2006; Bahlmann et al., 2009; Petersson et al., 2012; reviewed in Wilson et al., 2017). Specifically, it has been proposed that engagement of vlPFC follows a gradient of activation along the ventrodorsal axis that covaries with sequence complexity (Wilson et al., 2017; see Figure 1.2). Thus, there have been a number of efforts to describe the configuration of this area in terms of contributions to structural processing. Similar associations have been demonstrated between right inferior frontal gyrus and the processing of structure in music (Cheung et al., 2018; Koelsch et al., 2013).

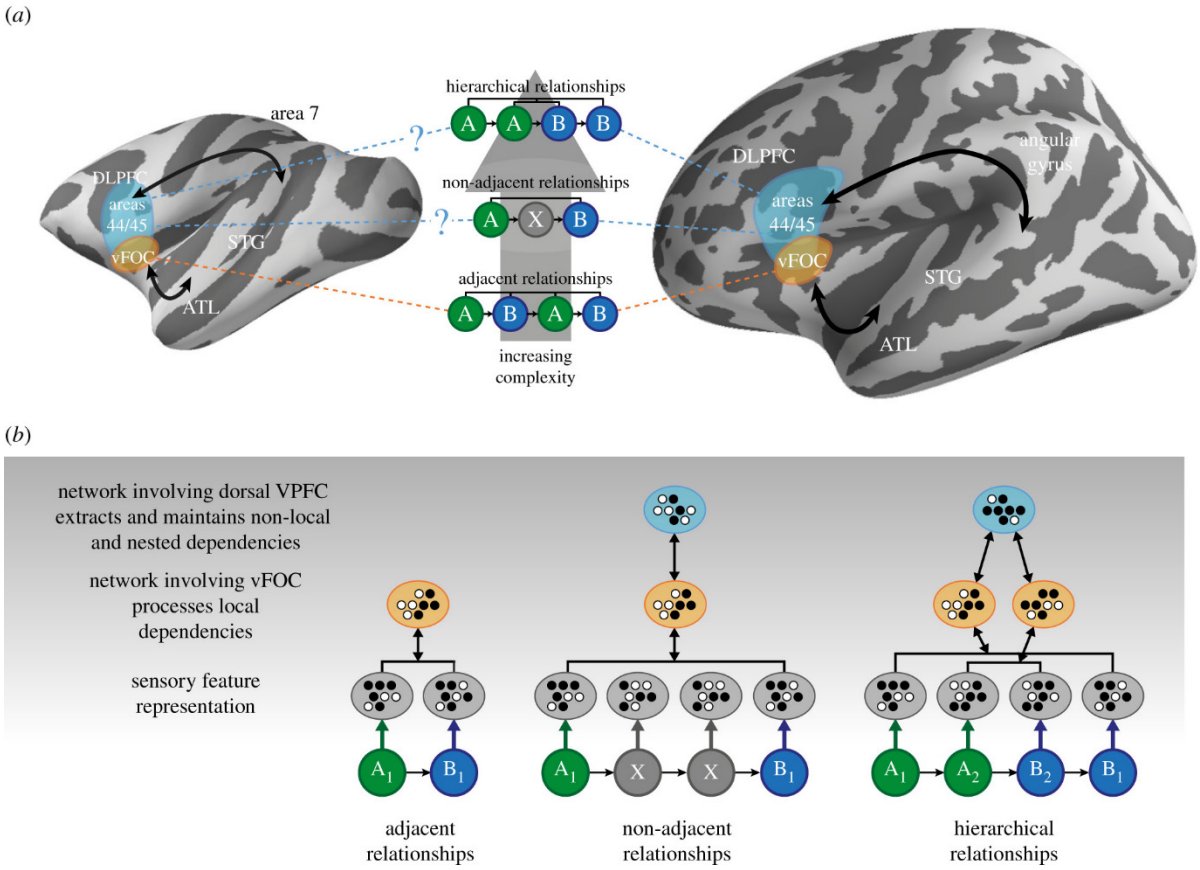


Figure 1.2 **Neurobiologically informed heuristic model of structured sequence processing, by Wilson, Marslen-Wilson & Petkov.** (A) Fronto-temporal regions involved in sequence processing, from Wilson et al., 2015 (Copyright © 2017 Benjamin Wilson, William D. Marslen-Wilson, and Christopher I. Petkov, CC BY 4.0). (B) Predicted combinatorial codes illustrated as neural patterns implemented by coordination between different regions.

Given the varied roles proposed for left inferior frontal gyrus (IFG), two pressing issues arise. Firstly, which, if any, of these putative functions truly reflects observed mechanistic contributions of IFG to sequence processing? And secondly, if – as might well be the case – all of these accounts have merit, how do we reconcile them? What might the neural codes in IFG actually represent?

In order to contribute to this area, we ideally require a task, such as an AGL task, that can elicit naturalistic structure-driven responses in vlPFC without eliciting confounding semantic responses. Furthermore, because responses to language unfold on the order of milliseconds (Beres, 2017), we require an imaging method capable of revealing responses in specific subregions of IFG, whilst

simultaneously revealing responses at a high time resolution. Functional magnetic resonance imaging (fMRI) has been used extensively to reveal localised sequencing-related activity at a high spatial resolution, revealing, for example, that a dorsal white matter tract appears to support complex syntactic processes, whilst ventral tracts subserve semantic processes and elementary syntax (Friederici & Gierhan, 2013). Conversely, electroencephalography (EEG) has been used to reveal syntax-related responses at a high time resolution, but with relatively poor spatial resolution (for example: Tabullo et al., 2013; Gunter et al., 2000; Hoen & Dominey, 2000; Hagoort et al., 1993). There is, however, a middle ground, in the form of intracranial electroencephalography, also known as electrocorticography (ECoG), which involves the surgical placement of electrodes within the cranium instead of upon the scalp, improving signal-to-noise ratio and spatial resolution relative to EEG (Asano et al., 2005). Moreover, such implantations typically also incorporate the surgical placement of depth electrodes that penetrate the cortical layers, thereby potentially providing access to hippocampal recordings as well, depending on clinical need. Because of the invasive nature of the procedure, however, human ECoG/depth electrode datasets are a relative rarity and can only be obtained in populations of willing patients undergoing clinically indicated intracranial monitoring, chiefly in cases of intractable epilepsy. However, through a cross-disciplinary transatlantic collaboration, we had access to a cohort of such patients (Chapter 3).

1.4 Sequences in mind: the neural representation of temporal structure

What is a sequence? It would not be incorrect to say: "A series of discrete elements, each following the last." However, from the perspective of the brain, it appears there are multiple answers to this question. Across a number of species, evidence suggests that there are multiple concurrent representations of sequence structure encoding orthogonal features within different brain networks (Dehaene et al., 2015). Below, I consider the evidence for various complimentary cognitive perspectives on sequence structure. This evidence suggests that multiple representations, instantiated concurrently by the human brain, contribute to structured sequence processing.

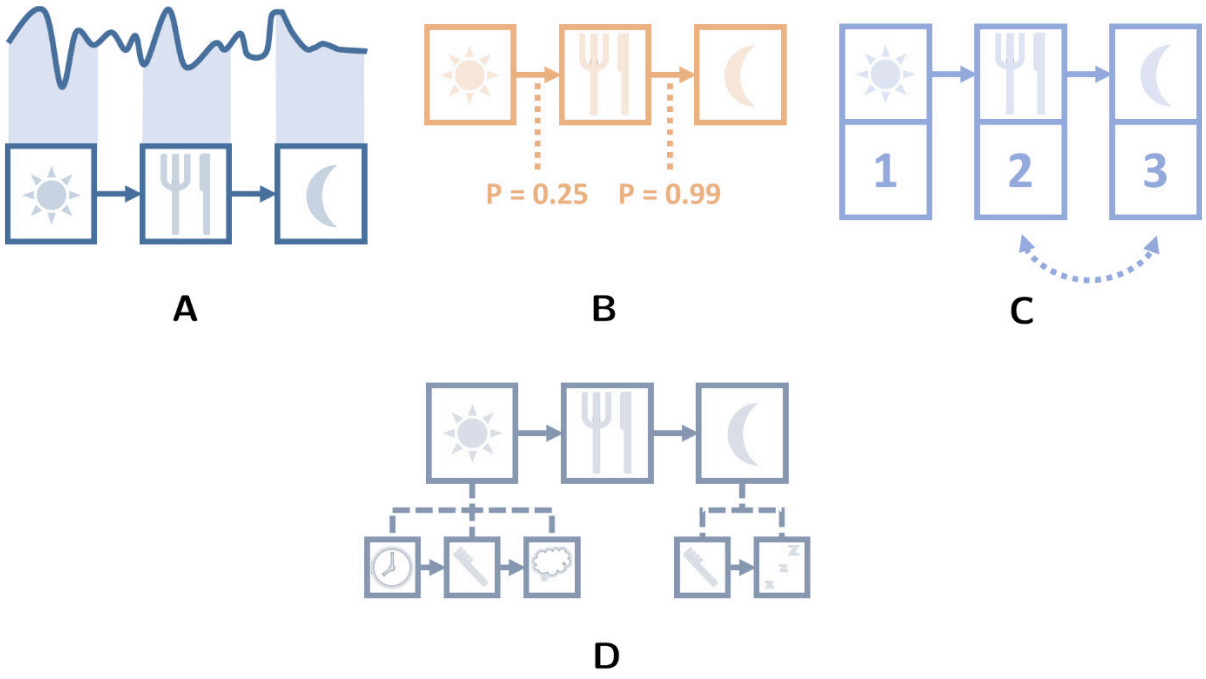


Figure 1.3: **A plurality of cognitive sequence representations.** A) Sensory stimuli are not inherently discrete, and must be discretised by segmentation processes. B) Segmentation both supports – and is supported by – transition probabilistic information, describing the likelihood of a transition between two specific discrete states. C) Items may be encoded on the basis of their ordinal position, which crucially simplifies the rearrangement of items by swapping their ordinal code. This rearrangement could be analogous to *syntactic movement*, the process by which perceptually separated items are mentally rearranged to support unification in language. D) Mental conceptions of sequences need not only consist of linear orderings of elementary items, but instead may also include tree structures, where subordinate sequences of items are hierarchically related to superordinate items.

1.4.1 Continuous signals, segmentation and discrete items

A speech stream is a continuously varying auditory signal, and yet humans are used to conceiving of discrete phonemes, words, phrases, sentences and sequences (Figure 1.3, A). However, moving from continuous to discrete representations is notionally non-trivial. From auditory input, the brain must undertake speech segmentation or "word discovery" to perceive discrete auditory units, which is all the more impressive given that pauses do not reliably mark the boundaries of words (Brent, 1999). In aphasic patients tasked with segmenting novel words of an artificial language, functional MRI (fMRI) localised phonological segmentation abilities to the left inferior frontal gyrus (IFG; Peñaloza et al., 2015). This is concordant with fMRI findings under a phonological judgment task suggesting that, with age, activation of dorsal IFG (BA 44/9) increases whilst

activation of superior temporal gyrus (STG) decreases, which the authors interpreted as reflecting a childhood shift in focus from auditory phonology to phonological segmentation (Bitan et al., 2007). However, IFG engagement in segmentation does not seem to be specific to speech alone. Similarly designed speech (Burton et al., 2000) and tone (Burton & Small, 2006) discrimination tasks have yielded fMRI evidence that left IFG is actually engaged in more general auditory segmentation. Furthermore, conjunction analysis of fMRI results has suggested that left IFG has a role in the segmentation of continuous stimuli across modalities, highlighting common patterns of activation in response to audiovisual presentation of novel Spanish and American Sign Language stimuli within a phonetic classification task (Williams et al., 2015).

The evidence is clear that segmentation is by no means a purely bottom-up, feed-forward process, but rather part of a feedback loop. For example, although prosodic and phonological cues can be used to delineate words in a continuous speech stream, these are not always sufficient in isolation. Human intracranial results appear to demonstrate that middle STG contains markers of stimulus-driven "syllable edge-detection" (Oganian & Chang, 2019), and behavioural evidence suggests that the influence of these speech cues may predominate over other factors in infancy (Johnson & Jusczyk, 2001). However, seminal studies in human infants, older children and adults have also suggested that continuous speech segmentation appears to be guided by the extremely rapid acquisition of statistical information on the co-occurrence of speech sounds (Saffran, Aslin, et al., 1996; Saffran et al., 1997). Here, participants were required to listen to a continuous stream of synthesised speech containing four meaningless tri-syllabic "words". Subsequently, they were assessed on their ability to discriminate these words from novel "non-words" containing identical syllables, but in an unfamiliar order. Participants of all ages exhibited a remarkable ability to establish word boundaries in a matter of minutes on the basis of the underlying statistical properties of the speech stream.

As we have seen, segmentation of continuous signals into discrete codes appears to be a vital part of the sequence learning process. However, more direct evidence has also been provided for the existence of discrete representations themselves. It has been proposed that, since action or policy selection produces discrete outcomes at specific times, discrete internal representations are required to support decision making (Parr & Friston, 2018). Most compellingly, a number of brain regions have been shown to instantiate what appear to be discrete encodings (for example of experiences, in

the hippocampus, see Shohamy & Wagner, 2008; of syllable onset events, in middle superior temporal gyrus, see Oganian & Chang, 2019; and of vowels and potentially auditory objects, in multiple regions, see Levy & Wilson, 2020 and Kuchibhotla & Bathellier, 2018).

1.4.2 Transition probabilities

The transition probability (or *transition probability*) of some state change **A→B** is the conditional probability of transitioning to state **B** given that state **A** has occurred. In sequence processing terms, the transition probability typically refers to the probability that some element **B** is upcoming given that **A** has just been encountered (note that this is actually the *forward transition probability*; the same principle, applied to items in reverse order, yields a *backward transition probability* that in some cases might prove more informative when resolving structure).

There is strong evidence that transition probabilities are of relevance to the learning of sequence structure. As previously described, continuous speech stream segmentation tasks have revealed that humans are able to utilise statistical information present within the stream to determine the boundaries of word-like units; specifically, the statistical information of relevance appeared to be the magnitude of transition probabilities (Saffran, Aslin, et al., 1996; Saffran et al., 1997). Moreover, electrophysiological findings have provided evidence that transition probabilities overtly influence neural responses and thus form part of the brain's repertoire of sequence representations (Figure 1.3, B). For example, oddball tasks have been found to elicit a characteristic electrophysiological response, the mismatch negativity (MMN), a broadly anterior response with an amplitude that, under controlled conditions, appears to be inversely related to transition probability (Koelsch et al., 2016). Additionally, recently reported findings have suggested that, during statistical learning of boundaries in continuous speech, transition probabilities manifest as influences on encodings within relatively early auditory processing regions including superior temporal gyrus (Henin et al., 2021). More generally, a theoretical relationship has been determined between the learning of transition probabilities and the formation of *chunks*, compressed and recoded aggregations of individual items that abstract away from needless detail to reduce cognitive load (Christiansen, 2019).

It is increasingly apparent that it is advisable to combine theories of transition probabilities with related theories of how exactly items are encoded in relation to each other. This is because, by encoding the relations of abstract or concrete items in a cognitive map, it becomes possible to infer

novel transitions that have yet to be encountered. This principle is known as *transitive inference*. Evidence suggests that inferential reasoning, including transitive inference, is supported by the hippocampus (Zeithamova et al., 2012; Zalesak & Heckers, 2009). Likewise, it is difficult to separate the concept of transition probabilities from the concept of *prediction* or *forecasting* of likely upcoming items. This, again, is presumed to fall within the remit of the hippocampus (Barron, Auksztulewicz, et al., 2020; Stachenfeld et al., 2017). As a consequence, it is becoming increasingly clear that well-triangulated computational models of hippocampal functioning have the potential to increase our understanding of sequence processing mechanisms (discussed further in 1.4.5).

1.4.3 Ordinal position, movement and restructuring

As we have seen, transition probabilities contain information integral to the discernment of structure in sequences. However, in many cases, transition probabilities are not the most efficient code by which to describe the salient features of a sequence. For example, the sequences **CABD**, **DAEF**, **XAYY** are most usefully described as "containing an A in the second position", rather than by reference to transitions. Encoding sequences by the numeric position of items is known as *ordinal positional encoding* (or simply *positional encoding*; Figure 1.3, C). This kind of encoding stands in stark contrast to the information provided by transition probabilities, because instead of describing pairwise associations between items (known in cognitive psychology as *associative chaining* or simply *chaining*), ordinal positional encodings allow the position of items to be described without reference to other elements of the sequence. A long history of behavioural study has suggested that associative chaining is unlikely to be the only method by which humans or their close relatives encode serially ordered items (Orlov et al., 2000; Restle & Brown, 1970; Lashley, 1951). More recently, discoveries have been made in non-human primates of ordinal-position-sensitive neurons in a number of regions, including primary motor cortex, premotor cortex, the supplementary motor area and intraparietal cortex (Carpenter et al., 2018; Nieder, 2012; Ninokura et al., 2004; Petrides, 1991).

1.4.4 Hierarchy

Hierarchical sequence structure (Figure 1.3, D), as we have mentioned, is a key structural feature of complex sequences in a variety of domains. Encoding of superordinate hierarchical structure in language has been localised by some accounts to the posterior portion of left Broca's area (BA 44; Friederici, 2020), and by others to the anterior portion (BA 45; Koechlin & Jubault, 2006).

More generally, equivalent roles have been proposed for right inferior frontal gyrus in the processing of hierarchical structure in music (Cheung et al., 2018).

Despite relatively consistent functional localisation, we still have a limited understanding of what exactly a "hierarchical" sequence encoding may look like in neurobiological terms. Temporally hierarchical codes, reflecting states at multiple levels of abstraction, have been found in the hippocampus (Takahashi, 2013), but it has been argued that these and other similar codes do not fully encompass the specific notion of sequential hierarchy relevant to syntactic processing, which involves storing a specific ordering of subordinate elements (Fitch & Martins, 2014). Even so, converging evidence suggests that the hippocampus supports inferences across levels of hierarchical structure (Whittington et al., 2020) and likely has a crucial role to play in hierarchial structure-building during sequence learning by way of interactions with prefrontal cortex (Theves et al., 2021; Opitz & Friederici, 2003).

Understanding hierarchical encodings likely requires a dynamic perspective. More specifically, in order to understand how the brain builds structure in general, we must propose specific dynamic mechanisms by which it is able to construct hierarchical neural codes through the recombination of constituents. In turn, this means we must ground any theoretical model of sequence processing in neurobiologically plausible solutions to the *neural binding problem* (Segaert et al., 2018; Feldman, 2013; Hagoort, 2005). Neural binding refers to the combination or association of neural signals to produce downstream effects, and it is posited to occur by many distinct processes. However, it has been suggested that specific formalisations of population coding and binding mechanisms, known as Vector Symbolic Architectures (VSAs), may shed light on these encodings and bridge the gap between abstract cognitive structure and neural implementations (Dehaene et al., 2015).

1.4.5 The need for computational unity

The ability to bridge the gap between cognitive representations and neural representations is important to the utility of a neurocomputational account of sequence processing. Marr (1982) famously defined three levels of description that are still widely applied in characterising any given model of the brain: the goals of a system (*computational* level), the cognitive processes required to reach these goals (*algorithmic* level), and the neural mechanisms required to instantiate them (*implementational* level). However, many existing perspectives on sequence processing inform

understanding at only a subset of these levels. For example, Chomsky (1956) famously established the importance of *formal language theory* as a computational account of grammatical complexity that echoes some of the sentiments expressed in Figure 1.1. More specifically, Chomsky defined a notional hierarchy of theoretical *automata* (the "Chomsky hierarchy") capable of recognising or generating specific types of structure (c.f. Figure 1.1, left versus right models, which have different capabilities). Much has been posited about human and animal computational abilities through the lens of the Chomsky hierarchy (Fitch & Friederici, 2012). However, such explanations support understanding primarily at the algorithmic and computational levels of Marr's framework. They do not, however, provide insights into how individual neurons or networks of neurons instantiate these algorithmic computations. Conversely, skyrocketing computational power in recent years has supported an array of approaches based on non-biologically modelled *artificial neural networks*, including *deep learning* models. These are, in effect, networks of interacting nodes connected by reconfigurable weights that can be tuned to "learn" patterns, including impressive sequence learning models (Alamia et al., 2020; Abiodun et al., 2018; Tosh & Ruxton, 2010; Moga & Gaussier, 2003).

Despite their remarkable ability to generalise from presented patterns, artificial neural networks do not generally incorporate any cognitively relevant algorithmic account for how their operations were accomplished. Arguably, the true way forward, therefore, is the coherent, large-scale simulation of artificial spiking neural systems, coordinated to support some broader algorithmically realised sequence learning goal, and thus spanning all of Marr's three levels. We have recently entered an age where it is possible to undertake such simulation at home, using freely available tools (Bekolay et al., 2014; Sherfey et al., 2018). As a consequence, we are now beginning to construct deeply informative models of site-specific mechanisms that abstract across Marr's levels, such as highly testable models of hippocampal function that may inform our understanding of sequence processing by triangulating a rich body of evidence (as reviewed in Hasselmo et al., 2020). For example, multiple models have now been designed that approximate the microcircuitry or cellular activity of the hippocampus whilst simultaneously accounting for its role in the sequential encoding and replay of episodic memory (for example: Sato & Yamaguchi, 2010; Galluppi et al., 2012; Trujillo & Eliasmith, 2014; Bayati et al., 2018). Even more impressively, Whittington et al. (2020) have produced a neurobiologically plausible hippocampal model that is capable of learning spatial and abstract relations, supports transitive inference, makes sensory predictions, and generates plausible activation

patterns for two major types of hippocampal cell, place cells and grid cells. Dumont & Eliasmith (2020) have similarly proposed a computationally specific correspondence between positional encodings and grid- and place-cell activity in the hippocampus.

Despite the impressive explanatory value provided by the models above, however, there remains a paucity of large-scale, neurobiologically plausible models incorporating representations of serially ordered sequences, especially those that can be generalised to explain functionality beyond the hippocampus (Choo & Eliasmith, 2010; Botvinick & Watanabe, 2007). Furthermore, no such models currently seek to attain more than superficial correspondence with site-specific neurobiological data across the frontotemporal language network. Development of such a model, able to inform our understanding of the neural codes, binding operations and cognitive representations relevant to sequence processing, is therefore a key goal of this schedule of work.

1.5 Conclusion

As this chapter has shown, both sequence learning tasks and studies of language, in conjunction with a host of broader neuroscientific findings, have helped us to refine hypotheses on the neurobiological substrates of sequence processing. However, despite this, we are still considerably lacking in our explanation of how exactly we process sequences of items separated over time. We lack strong evidence supporting specific accounts of how – or even how well – the brain learns particular statistical regularities between items; how it codes representations of sequential items in key areas of the frontotemporal network; and how such representations might be combined or restructured into new forms. Thus, the primary aim of this schedule of work was to clarify human sequence learning capabilities and explore their neural substrates. Additionally, we aimed to integrate new and existing findings into a coherent and internally consistent neurocomputational account of sequence processing, and to develop novel analyses to support the testing of this model against neural data. Although we made use of proto-linguistic stimuli in pursuing these aims (see Chapters 2 and 3), it is important to note that our primary interest was not in language itself, but rather in the use of naturalistic stimuli to provide a window on more generic sequence processing mechanisms (see 1.2).

In this thesis, I present behavioural, electrophysiological and computational evidence directed towards these goals. In Chapter 2, I report the behavioural results of an AGL experiment characterising adult human abilities to learn adjacent and non-adjacent dependencies concurrently. Here, I show that the cohort learned both types of dependency. I also present evidence that implicit reaction times, rather than seemingly explicit grammaticality judgement measures, are a more sensitive measure of artificial grammar learning. In Chapter 3, I analyse behavioural and electrophysiological results collected under this AGL paradigm in neurosurgical patients, employing a group-level univariate approach to show that ECoG and depth electrode responses to adjacent and non-adjacent dependencies included significant effects within key regions of a frontotemporal network including IFG pars opercularis and the hippocampus. However, I also suggest that the results argue for the use of novel multivariate approaches to advance our understanding beyond the insights afforded by traditional methods. In Chapter 4, I integrate cutting-edge and novel computational approaches with a wide-range of existing neurobiological evidence, triangulating them to produce a neurobiologically plausible computational model of sequence processing in a number of regions, *VS-BIND* (Vector-symbolic Sequencing of Binding INstantiating

Dependencies). Here, I show that VS-BIND generates a host of site-specific predictions at multiple levels of Marr's framework, ripe for testing. In Chapter 5, I present the results of novel and existing multivariate analyses on the intracranial AGL task data, first demonstrating alignment of results under this approach with those of more traditional functional connectivity analyses; then providing evidence for key sequence representations posited in VS-BIND to exist in motor cortex and inferior frontal gyrus; and ultimately revealing inter-regional flow of representations potentially consistent with the predictive coding framework (Summerfield & de Lange, 2014; Friston & Kiebel, 2009; Rao & Ballard, 1999), which suggests that context-driven predictions of input from higher-order regions interact with input from lower-order regions to produce prediction errors. Finally, in Chapter 6, I conclude the thesis by discussing the implications of these combined results. I consider how this research has informed understanding of human sequence processing abilities, and how it informs our understanding of the specific roles undertaken by key regions of the frontotemporal network during sequence processing. I lastly discuss exciting avenues proceeding from this research, and which have the potential to further advance neurocomputational accounts of language and cognition in general.

Chapter 2. Behavioural assessment of mixed-dependency artificial grammar learning

All work presented in this chapter was primarily conducted by the author, except where explicitly stated. The author designed and implemented the behavioural task in MATLAB following discussion with academic supervisors. Behavioural data were collected with Oana Morteanu, an undergraduate project student, with support from Benjamin Wilson. The overwhelming majority of analysis and writing was conducted by the author. Supervisors provided valuable comments and discussion.

2.1 Abstract

Adjacent dependencies (ADs) are relationships between neighbouring sequential items and occur abundantly in nature and language. Non-adjacent dependencies (NADs) are also omnipresent in language and a precursor to hierarchical dependencies, which include the nested relationships that imbue language with infinitely expressive capabilities.

Dependency learning can be studied in relative isolation, without the confounds of natural language. Artificial Grammar Learning (AGL) tasks expose participants to sequences conforming to a covert set of ordering rules (a grammar), and subsequently assess participants on their ability to distinguish novel "grammatical" sequences from "ungrammatical" sequences, for example by way of a grammaticality judgement task. AD and NAD learning have both been studied with AGL tasks, but there is a lack of evidence directly comparing NAD and AD learning under a single task. Evidence suggests NAD learning is cognitively demanding in specific cases, but human dependency learning abilities need clarification. Additionally, studies showing NAD learning often overtly distinguish salient from non-salient items by their phonological properties. However, as a model of natural language syntax this constraint lacks ecological validity.

To simultaneously assess human AD- and NAD-learning abilities using appropriate cues, I devised a novel auditory AGL task using sequences containing either ADs or NADs. This task incorporated cues including prosody, feedback, delineated sequence boundaries, staged pre-exposure, and variable intervening items. Across 50 healthy adults, a grammaticality judgement measure revealed above-chance performance. However, further analysis revealed bimodal performance despite all cues, with some participants performing at very high levels and others at chance, suggesting the grammaticality judgement process is cognitively non-trivial. Despite this, even in chance performers, reaction time measures revealed differences in responses to grammatical and ungrammatical sequences, suggesting implicit learning even in those who failed to demonstrate explicit knowledge of the grammar.

This study replicates findings that humans can learn NADs without the benefit of phonological cues, but extends them by providing insights into performance variability under AGL tasks, suggesting that implicit, rather than explicit, measures may better reflect grammatical sensitivity. Finally, this task forms the basis for the electrophysiological study of sequence processing reported in Chapter 3 and Chapter 5.

2.2 Introduction

In the previous chapter, I argued that there is a pressing need to better characterise the mechanisms by which the brain is able to reconcile information distributed across time. This characterisation naturally begins at the level of behaviour. In this chapter, I report the findings of a novel artificial grammar learning (AGL) experiment designed to measure sensitivity to relationships between stimuli that are either adjacent in time, or more distantly separated.

As we have seen, the ability to extract and reconcile events distributed in time is critical to many domains including music, mathematics, action sequencing and general cognition (see Chapter 1). The limits of these capabilities, however, are exemplified in humans by our facility with natural language. Natural language features complex local and nonlocal dependencies between phonemes, morphemes, words, phrases and beyond. These dependencies must be rapidly consumed and resolved in order to parse meaning during language comprehension and, conversely, instantiated as part of a linear stream of fluent utterances during speech production. Language therefore offers an enticing route to an improved neurobiological understanding of human structured sequence processing.

Natural language offers examples of adjacent and non-adjacent dependencies at every structural level. At the morphemic level, for example, English features prominent relationships between word stems and inflectional morphemes. These include the adjacent relationship between the stem of a regular verb and "-ed" to conjugate the past tense. They also include the somewhat more complex non-adjacent relationship between an auxiliary verb and the suffix "-ing" (as in "is X-ing", in which the intervening verb stem is largely uninformative from a syntactic perspective; Friederici et al., 2011). At the level of words, adjacent statistical dependencies exist between adjectives and the neighbouring nouns or noun phrases they commonly modify (for example, "blue cheese"), and non-adjacent dependencies exist between modal verbs and the main verb in question form (as in "Can you see..?" or "Would they take..?"). Relationships spanning phrases are ubiquitous, too, as in the non-adjacent example of subject-verb agreement (for example, spanning the noun and verb phrases of "the PhD student upstairs studiously writes" versus "the PhD students upstairs studiously write"; see Wilson et al., 2018). Furthermore, not only are adjacent and non-adjacent dependencies frequently both present between the tokens of a single utterance at all of these levels, but it is also

possible to find many examples where identical or similar tokens can legally form both adjacent and non-adjacent dependencies in different scenarios (Deocampo et al., 2019).

Finally, also at the phrasal level, multiple dependencies are commonly related in a hierarchical manner, for example through nesting or tail recursion, or a mixture of both, as in "The dog, [who chased the cat, [who caught the mouse]], barked" (Rohrmeier et al., 2012). A similar form of hierarchical structure, cross-serial dependencies (of the form "$\underline{A_1}\ A_2\ \underline{B_1}\ B_2$") is present in Dutch and Swiss German (Christiansen & MacDonald, 2009). Although the degree to which recursion is a defining characteristic of human language, and the practical importance of such claims, is disputed (Hauser et al., 2002; Jackendoff & Pinker, 2005; Pullum & Scholz, 2010), any grammar featuring recursive constructions possesses theoretically unbounded expressivity at a structural level (Harris, 1957). The ability to process multiple concurrent non-adjacent dependencies is a prerequisite for both nested and cross-serial dependency comprehension and, as such, non-adjacent dependencies support major aspects of natural language recursivity.

In summary, adjacent and non-adjacent dependencies are ubiquitous in natural language and essential to its communicative properties, and the behavioural and neurobiological study of natural language must therefore inevitably inform accounts of sequence processing. However, language incorporates not only structured organisation, or *syntax*, but also a lexicon of referents describing real-world percepts or concepts, and these imbue our communications with interpretable meaning, or *semantics*. Furthermore, natural language comprehension relies, at the one extreme, upon the interpretation of broad contextual cues and, at the other, on well-learned patterns of individual phonemes, the subjects of the fields of *pragmatics* and *phonology*, respectively. Whilst language syntax is a striking example of human dependency processing capabilities, as an example of sequence processing, natural language is thus inherently contaminated by pre-learned phonology and pragmatics, as well as by the confounds of semantics, the processing of which engages many regions, including some that are spatially coincident with syntax-sensitive areas (Petersson et al., 2012; Petkov & Wilson, 2012; Fuji et al., 2016).

Artificial grammar learning (AGL) tasks are one method by which dependency learning can be studied in relative isolation from other aspects of language. AGL tasks are a subset of artificial language learning tasks (Ettlinger et al., 2016) that involve first exposing participants to sequences

conforming to an implicit set of ordering rules, known as an *artificial grammar*, and subsequently assessing participants on their knowledge of the grammar, for example on the basis of their ability to actively distinguish novel "grammatical" sequences from "ungrammatical" sequences (ten Cate et al., 2020). If, after exposure to the training stimuli, the subject is able to detect the rules or statistical regularities that characterise the AG, the subject should exhibit measurable sensitivity to the grammaticality of the test stimuli. Remarkably, humans, including infants, can learn to identify these characteristic regularities following as little as two minutes of incidental exposure (see, for example, Saffran, Aslin, et al., 1996), although the matter of what exactly is learned during these tasks is still a matter of some debate (see Alhama & Zuidema, 2019; Marcus et al., 1999) even sixty years after the earliest AGL tasks were devised (Miller, 1958; Reber, 1967). There is also strong evidence in humans that AGL tasks engage natural-language-relevant neurobiological processes (Friederici & Chomsky, 2017; Ettlinger et al., 2016; Tettamanti & Perani, 2012; Romberg & Saffran, 2010; Petersson et al., 2004). Moreover, performance on AGL tasks is correlated with natural language abilities (Frost et al., 2015), and impacted in cases of non-fluent aphasia (Cope et al., 2017; Conway & Pisoni, 2008; as reviewed in Wilson et al., 2017).

By avoiding semantic prerequisites, AGL tasks offer a "relatively uncontaminated window" into the neurobiology of dependency processing in human adults (Petersson et al., 2012). Likewise, whilst some of the earliest AGL tasks, such as that of Reber (1967), used written verbal stimuli in humans, the flexibility of the paradigm means that AGL tasks have subsequently been employed to study responses to verbal and nonverbal stimuli in multiple sensory domains and species, for example using abstract visual stimuli (Stobbe et al., 2012), auditory speech and tone stimuli (Saffran, Aslin, et al., 1996; Saffran et al., 1999), motoric stimuli (Opacic et al., 2009) and olfactory stimuli (Shahbaba et al., 2019).

2.2.1 Adjacent and non-adjacent dependency learning abilities in humans

Human adults, children and infants are capable of very quickly learning to identify word boundaries in rapidly unfolding speech streams, seemingly grouping adjacent syllables into word-like units using the information described by transition probabilities (Newport & Aslin, 2004). In contrast, although it is clear from the prevalence of non-adjacent dependencies (NADs) in language that they are critically important and that humans are able to process them, AGL studies of NAD learning have yielded inconsistent results. It has been suggested that this inconsistency arises because

the ability to learn NADs is far more selective compared to adjacent dependency learning (Gebhart et al., 2009).

It is certainly true that humans can learn NADs outside of language. A number of statistical learning studies provide evidence of this under specific circumstances. Studies that have been designed specifically to test simple NAD learning have shown that humans are able to respond to violations of learned relationships between non-adjacent elements (Watson et al., 2020; Deocampo et al., 2019; Vuong et al., 2016; Romberg & Saffran, 2010; Gebhart et al., 2009; Lany & Gómez, 2008; Newport & Aslin, 2004; Gómez, 2002; Peña et al., 2002). Statistical learning studies designed to demonstrate learning of hierarchical dependencies (e.g. centre-embedded or "nested" dependencies) also provide evidence of NAD learning outwith language, since reconciliation of non-neighbouring items is a necessary prerequisite for the encoding of multiple nested relationships (for example, Friederici, Fiebach, et al., 2006; see de Vries et al., 2011 and Wilson et al., 2018).

Given the above, it might be concluded that humans are trivially capable of learning NADs in natural and artificial languages. However, it appears that NADs are difficult to detect except under specific circumstances. For example, Newport & Aslin (2004) determined, by running a series of increasingly simplified statistical learning tasks in adult participants, that subjects could not exceed chance performance in differentiating grammatical from ungrammatical non-adjacent stimuli except where the critical stimulus elements were distinguished from each other at the perceptual level (e.g. **a<u>X</u>b** or **<u>A</u>x<u>B</u>**). Similarly, Onnis et al. (2005) determined that, when processing a novel AG, extraction of NADs from continuous speech was possible based on distributional properties, but only when dependent items could be grouped by phonological similarity.

2.2.2 Ecologically valid aids to artificial grammar learning

It may seem contradictory that multiple AGL studies have been able to demonstrate significant NAD learning, given that such dependencies are apparently challenging to detect except in specific circumstances, for example when supported by phonological cues. In fact, it is often the case that auditory AGL studies include sequences where the dependent and irrelevant items are grouped into distinct phonological categories, and usually as a deliberate design choice. These categories are namely based on segmental phonology (i.e. the combination of phonemes in each item). For example, the sequence elements used by Peña et al. (2002), who revealed learning of NADs

in a human AGL task, were presented as a continuous stream, but could be separated into "dependent" or "irrelevant" on the basis of whether they contained either plosive (*p*, *k*, *b*, *d*, *t* or *g*) or continuant (*l*, *r*, *f*) sounds. The elements used by Gómez (2002), which have been incorporated into a number of subsequent studies, were of different syllabic lengths and separated by pauses during presentation (e.g., 'po gafo di', where the dependency is between 'po' and 'di', but 'gafo' is an irrelevant intervening element). Elements could thus be separated into "dependent" or "irrelevant" on the basis of whether they were monosyllabic or disyllabic, respectively (although this does not detract from the importance of this seminal study; see below). Although this kind of role-specific phonological grouping is likely more salient in a small artificial vocabulary than it is in natural language, it has been suggested that such cues are ecologically valid because, intriguingly, words from the same grammatical category tend to be phonologically coherent in language (Monaghan et al., 2007). Therefore, groupings based on segmental phonology are not only prevalent in the AGL literature and highly beneficial to NAD learning, but also relevant to the learning of dependencies in natural language. However, segmental phonological cues are by no means the only acoustic features to aid learning in auditory AGL tasks. Prosody is an essential component of natural speech, encompassing non-segmental phonological features such as intonation that serve to emphasise particular words or phrasal boundaries. Prosodic cues that have been shown to assist learning of NADs include pitch and intensity, where emphasis of either the dependent or irrelevant items is sufficient to aid learning (Grama et al., 2016), and pauses in or around sequences (Hawthorne & Gerken, 2014).

As well as phonological cues, statistical manipulations can also be assistive to NAD learning. In a seminal study by Gómez (2002), adults and 18-month-old infants were exposed to 3-element auditory sequences of nonsense words spoken by an adult female. Sequences followed the form **AXB**, comprising a deterministic NAD in which the dependents **A** and **B** were separated by a single intervening item, **X**. All items and sequences were separated by pauses. The variability of **X** was manipulated between subjects by varying the set size from which **X** was chosen. In adults, learning was then assessed by a grammaticality judgement task, and in infants, by using an orienting response (head-turn preference procedure, see Kemler Nelson et al., 1995) to test discrimination of grammatical versus ungrammatical sequences. In the highest variability condition, where **X** was shown to take one of 24 possible values during training, adults and infants exhibited significantly

greater sensitivity to the grammaticality of the NAD during testing than under the lower variability conditions. Thus, it was found that learning of NADs only occurred once any ADs were rendered sufficiently unpredictable by the presence of a highly variable intervening element, suggesting that the reliable, deterministic NADs were prioritised over the unreliable, probabilistic ADs.

In light of these and similar results, it has been hypothesised that the dependency learning mechanism prioritises those dependencies that yield the most reliable information (Onnis et al., 2004). This hypothesis similarly predicts that nondeterministic NADs should be very difficult to learn. Consistent with this account, a study by van den Bos et al. (2012) showed that adults could learn deterministic NADs, but could not learn probabilistic NADs without the aid of additional sensory cues. This was despite the fact that the intervening and dependent items were distinguished by differing syllabic lengths, and that items and sequences were surrounded by pauses.

Finally, in addition to phonological and statistical cues, extra-linguistic factors can aid artificial grammar learning. For example, learning can be facilitated by a "starting small" approach (Elman, 1993), which refers to incremental training in which simple relationships are presented before complex ones. This is thought to mimic incremental exposure to natural language, and has been shown to enhance learning in complex artificial grammar tasks; participants first presented with ADs, for example, which are easier to recognise and learn than NADs, are better able to subsequently generalise these to more complex NADs (Lany et al., 2007; Lany & Gómez, 2008; Lai & Poletiek, 2011; Poletiek et al., 2018).

In summary, by comparison with AD learning, AGL studies of human NAD learning have yielded inconsistent results that require clarification. Furthermore, the inconsistencies in these results suggest a high degree of sensitivity to specific task parameters, with learning benefitting from conditions that bring the dependents "closer together ... physically, attentionally, or perceptually" (Wilson et al., 2018), including phonological cues, iteratively ordered learning, and manipulations of item variability. However, very few studies have studied the concurrent learning of AD and NAD dependencies in humans or other primates in order to allow controlled comparisons to be made between them (Deocampo et al., 2019; Vuong et al., 2016; Wilson, Smith, et al., 2015; Romberg & Saffran, 2013). Of these named studies, only Deocampo et al. (2019), Vuong et al. (2016) and

Romberg & Saffran (2013) comprised statistical learning tasks run in humans, and only the latter two of these featured naturalistic speech stimuli.

2.2.3 Aims

The primary aim of this experiment was to assess the ability of healthy adult participants to learn a novel artificial grammar containing both adjacent and non-adjacent dependencies, with the goal of using this task in a clinical population. In addition, we used three different manipulations of the prosody of the sequences to assess whether additional phonological cues to the dependencies, present either during exposure or throughout the whole task, would increase performance over a "no-prosody" condition.

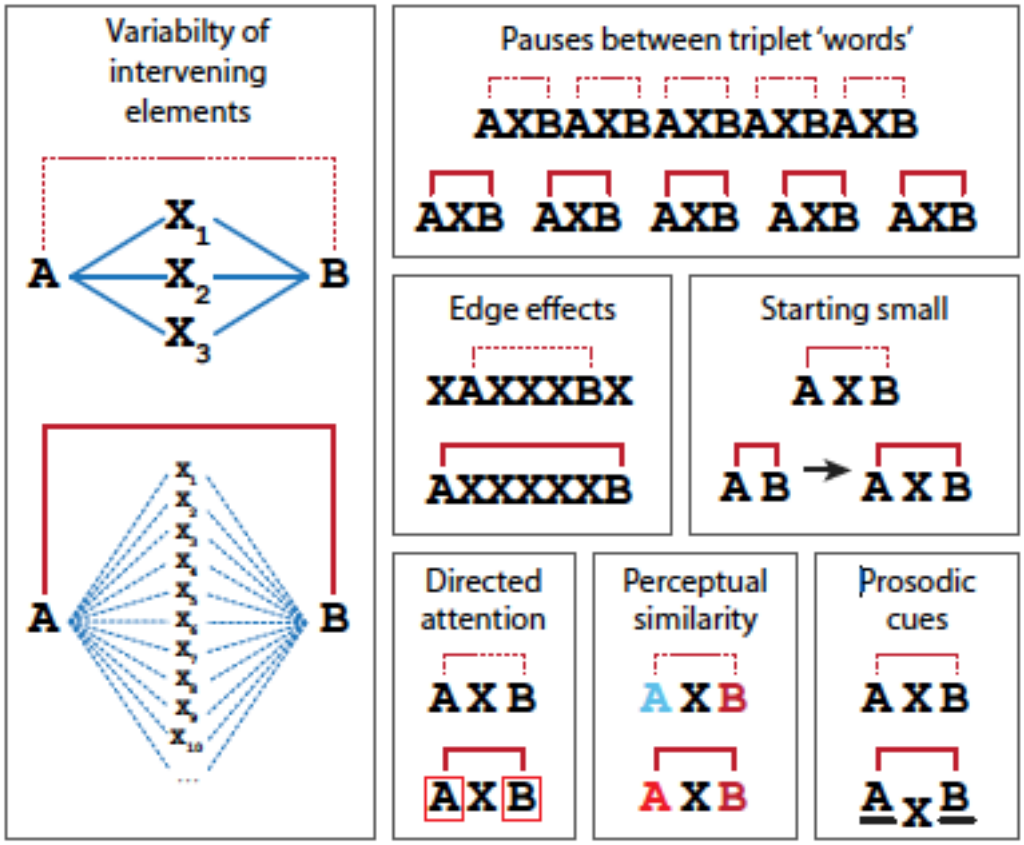


Figure 2.1: **Cues that aid non-adjacent dependency learning (see Wilson et al., 2018; *reused with permission*)**. In developing a task that can be used to assess AD and NAD learning in neurosurgery patients, we made use of many of these, including variability, pauses, edge effects, starting small and prosodic cues.

To achieve this, I developed an AGL task that included both ADs and NADs, allowing within-subject comparisons of performance on different dependency types. The sequences consisted

of monosyllabic, auditory nonsense words. A number of aids to learning were incorporated (see Figure 2.1), including: pauses around individual sequences; highly variable intervening elements; feedback; a limited number of relationships; and the generalisation of the pair relationship to both AD and NAD cases; but expressly excluded the labelling of item roles by overt perceptual cues, other than the prosodic manipulations we applied. We also used a "starting small" approach to aid the simultaneous learning of ADs and NADs, and to facilitate learning in general. We hypothesised that, through limited (pre-)exposure to the grammatical sequence set, participants would be able to learn non-adjacent dependencies with the benefit of these aids to learning, and that the presence of prosodic cues during exposure and/or testing would further enhance performance.

This has the potential to extend the literature on human dependency learning abilities, adds to the currently limited existing evidence on simultaneous AD and NAD learning, and serves as a crucial precursor to a future electrophysiological task. From this perspective, it is important to behaviourally characterise sequence learning abilities and strategies employed under the task in order to support the interpretation of future neurobiological findings.

2.3 Methods

2.3.1 Participant recruitment and ethics

Participants (n = 50) comprised a mix of registered students and members of the wider public, recruited for behavioural study at Newcastle University via the Department of Psychology's Research Participation Scheme (students) and the Institute of Neuroscience Participant Pool (non-students). To meet inclusion criteria, participants were required to have normal hearing, normal or appropriately corrected vision, not to have participated in any prior sequence learning study, and to be aged 18 to 60 (μ = 23, σ = 7.42 years, range 18–51). Student participants were compensated for their time with course credits, whilst non-students received a £5 Amazon voucher. All were informed that they were free to leave at any time without giving any reason. Before undertaking the task, each participant read and signed a consent form. Following completion of the behavioural task, each participant was asked to complete a questionnaire, described below (see also Appendix 3: Behavioural questionnaire). Full ethical approval for the schedule of research was sought and granted by the Faculty of Medical Sciences Ethical Committee.

2.3.2 Task design

The *AxB* task was an auditory Artificial Grammar Learning (AGL) task running on a Dell desktop computer and coded in Psychophysics Toolbox (Kleiner et al., 2007) using MATLAB (The MathWorks, Inc.). The task was conducted within a psychophysics laboratory. Auditory stimuli were presented at approximately 75 dB using circumaural headphones (Sennheiser HD 202) and visual information was displayed on a 24 inch monitor. Participants were seated approximately 60 centimetres from the screen and used a computer keyboard to interact when required. Whenever onscreen instructions were presented, participants could iterate through these at their own pace using the keyboard.

Before undertaking the task, each participant was advised that their first objective was to pay attention to the relationships between the words they would hear. No further information was provided on the intended structure or vocabulary of the stimuli. Such minimal instructions are typical of AGL tasks in human adults. In particular, explicit instruction is avoided in order to promote implicit statistical learning of the underlying sequence structure (Schiff et al., 2017).

The task itself lasted approximately 30 minutes and was divided into three consecutive phases: a passive *pre-exposure* phase; an *AG exposure* phase (which corresponded to the traditional starting point for an AGL task, during which stimuli fitting a covert grammar were presented to the participant); and an active *testing* phase. Figure 2.2 illustrates the task. The *AG exposure* phase followed the canonical arrangement of presenting stimuli drawn only from the underlying artificial grammar (see ten Cate et al., 2020), and is described below. Understanding the AG exposure and test phases is a prerequisite for understanding the motivations behind the pre-exposure phase, the subject of which will therefore be returned to later in the text.

AG exposure comprised 48 trials in which the participant was presented with a sequence that followed the rules of an underlying artificial grammar (see below; also see Table 2.2). Each trial required only passive listening to the auditory sequence. To match the conditions required of a prospective electrophysiological experiment and to prepare participants for the requirements of the testing phase (see below), participants were requested to stare at a white fixation cross in the centre of the screen during this phase (Figure 2.2, **1**). Sequences are described below in more detail but each

lasted 2100 ms and, during exposure, were separated by silent inter-trial intervals of 2150 ms, yielding an exposure phase of just under 4 minutes in length for all 48 trials.

The active testing phase consisted of two runs of two-alternative forced choice testing, each of which consisted of 48 trials and lasted roughly 6 minutes. At the beginning of each testing run, participants were informed through onscreen instructions that some of the sequences they were about to hear would contain the important relationships they had already heard, and some would not.

In each trial of the testing phase, participants were presented with a novel sequence that was either grammatical or ungrammatical, and asked to categorise it as either "fitting the pattern" of the sequences they had previously heard, or "not fitting the pattern" (instructions after Milne et al., 2018). Participants responded by pressing one of two buttons ("C" and "M" keys, counterbalanced between participants). Trials were presented in a randomised order, with 50% of trials (i.e. 24 trials per run) being grammatical, and 50% ungrammatical. Trials did not time out, and so individual testing time varied by reaction speed. Throughout playback of each auditory sequence, a yellow fixation cross was shown in the centre of the screen (Figure 2.2, **2**). To discourage explicit cognitive strategies in the absence of any time limit, participants were asked to respond swiftly to the grammaticality judgement task based on their "gut feeling", choosing the appropriate keyboard response ("C"/"M") as soon as the central fixation cross turned from yellow to blue (Figure 2.2, **3**). Each sequence lasted 2100 ms, followed by a silent, jittered pre-response period of between 1000 and 1300 ms, before the fixation cross turned blue. Following a response, the computer provided the participant with onscreen feedback by briefly changing the colour and/or orientation of the fixation cross (a red "X" for incorrect judgements, and a green "+" for correct judgements; Figure 2.2, $\mathbf{4_A}$ and $\mathbf{4_B}$). A jittered post-response delay of between 1000 and 1300 ms followed each response.

Between the two 48-trial testing runs, participants experienced a 2 minute refamiliarisation period, which was identical to the exposure phase except for its shorter duration. Refamiliarisation between testing runs is common in AGL tasks and intended to mitigate the effects of the ungrammatical sequences that form 50% of the test set, which might otherwise induce unwanted interference with the learning that took place in the exposure phase (Kuppuraj et al., 2018).

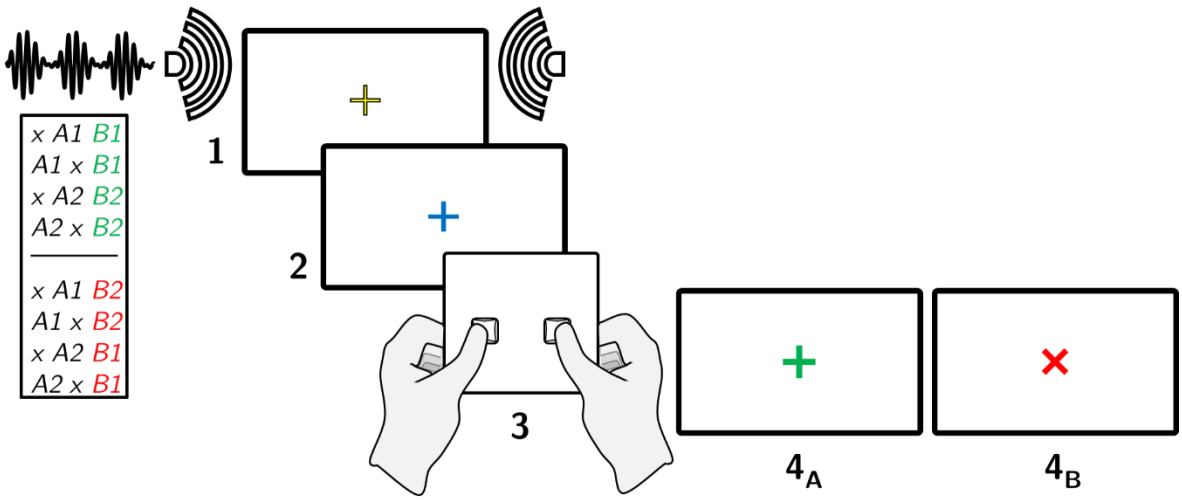


Figure 2.2 ***AxB* task diagram**. During exposure, participants listen to a set of valid stimulus sequences (terminating in green, left) whilst looking at a fixation cross (1). During testing, participants do the same but hear a 50-50 mix of valid and invalid stimulus sequences (terminating in red) whilst waiting for a fixation cross to change colour from yellow to blue (2). Once cued, participants respond using keyboard keys (3) and receive feedback on their correct (4A) or incorrect (4B) response.

2.3.3 Stimuli

Stimuli were auditory sequences comprised of three naturally spoken, single-syllable nonsense words, based on stimuli designed and used in previous AGL studies (Wilson, Smith, et al., 2015; Saffran et al., 2008; Saffran, 2002). Nonsense words were used as they maintain the spectrotemporal complexity of real words, and are readily distinguishable (Wilson, Smith, et al., 2015). Each word followed a consonant-vowel-consonant construction and was designed to avoid combinations considered common and meaningful in English.

Nonsense words had previously been produced by an adult female speaker in monotone and recorded using an Edirol R-09HR (Roland Corp.) sound recorder. Recorded words were stored in digital sound files lasting precisely 600 ms each, then root-mean-square balanced and combined into three-word sequences using a custom MATLAB script that inserted 150 ms inter-stimulus intervals (ISI), producing a set of three-word stimuli each lasting 2100 ms. This constant playback time ensured that neither the duration of individual words nor the duration of the sequence as a whole could be used as cues to grammaticality. The complete list of words can be found in Table 2.1 below, along with their transcription in the International Phonetic Alphabet (IPA) for avoidance of ambiguity.

To assess the possible benefit of prosodic pitch cues on performance in this task, across the cohort, participants were randomly assigned to one of 3 prosodic manipulation groups before the task began: NP ('no prosody'), PE ('prosody during exposure only') and PT ('prosody during both exposure and testing'). Prosodic trials were identical to non-prosodic trials except that, for prosodic trials, the frequency content of **X** elements was pitch-shifted down by 5% relative to the fundamental frequency baseline (using the Praat software package; Boersma & Weenink, 2016), and the frequency content of **A** and **B** elements was pitch-shifted up by 5% relative to baseline. This produced a net 10% pitch difference between **A**/**B** and **X** words, designed to emphasise the distinction between these elements. This was considerably larger than the "just noticeable difference" reported in the psychophysical literature for speech pitch manipulations (Liu, 2013), and resulted in naturalistic utterances. Sound amplitude and duration were not varied.

Table 2.1: **Nonsense words used in the AxB task, and IPA transcriptions**

Label	Word	IPA	Label	Word	IPA
X_1	Bek	bɛk	X_{17}	Pif	pɪf
X_2	Ber	bɛr	X_{18}	Pux	pʌks
X_3	Biz	bɪz	X_{19}	Raz	ræz
X_4	Dil	dɪl	X_{20}	Rik	rɪk
X_5	Fal	fæl	X_{21}	Rit	rɪt
X_6	Gol	gɒl	X_{22}	Ruj	rʌʤ
X_7	Heb	hɛb	X_{23}	Yun	jʌn
X_8	Hig	hɪg	X_{24}	Zil	zɪl
X_9	Hok	hɒk			
X_{10}	Jat	ʤæt			
X_{11}	Kay	keɪ	A_1	Fip	fɪp
X_{12}	Kiv	kɪv	A_2	Gak	gæk
X_{13}	Kug	kʌg	B_1	Wez	wɛz
X_{14}	Lar	lɑː	B_2	Lod	lɒd
X_{15}	Lun	lʌn			
X_{16}	Mot	mɒt			

All stimuli were generated according to the following rules. Every stimulus sequence was generated from a mixed dependency grammar permitting *either* an adjacent dependency *or* a non-adjacent dependency within each sequence. "Adjacent" sequences followed the form **xAB**, where **A** and **B** represented the important constituent words of the dependency, but the identity of **X** was irrelevant to the relationship under test. "Non-adjacent" sequences followed the form **AxB**, where again **A** and **B** represented the critical components of the dependency, and the intervening element **X** again denoted any word irrelevant to this relationship. In order to highlight the importance of **A** and **B** whilst diminishing the salience of **X**, there were 24 phonetically distinct **X** items but only 2 **A** and 2 **B** items (see Table 2.1). As described in 2.2.2, previous research in human adults and infants has indicated that deliberately varying irrelevant items in this way can reduce their saliency within an

artificial grammar and promote the learning of non-adjacent dependencies (see especially Gómez, 2002).

Whilst grammatical sequences and ungrammatical sequences incorporated the same vocabulary of nonsense words, grammatical test sequences were distinct from ungrammatical test sequences in terms of the rules they followed. Grammatical sequences followed either of the patterns $A_1 \rightarrow B_1$ or $A_2 \rightarrow B_2$, where "→" denotes the relationship "is ultimately followed by". For $A_1 \rightarrow B_1$, for example, this encompassed the following valid sequences: xA_1B_1 and A_1xB_1. Likewise, for $A_2 \rightarrow B_2$, valid sequences included xA_2B_2 and A_2xB_2. Note that every sequence ends in a **B**. Thus, there were two possible valid **A→B** pairings, $A_1 \rightarrow B_1$ and $A_2 \rightarrow B_2$, and these could both occur in either adjacent or non-adjacent configurations. This grammar is illustrated in Figure 2.3.

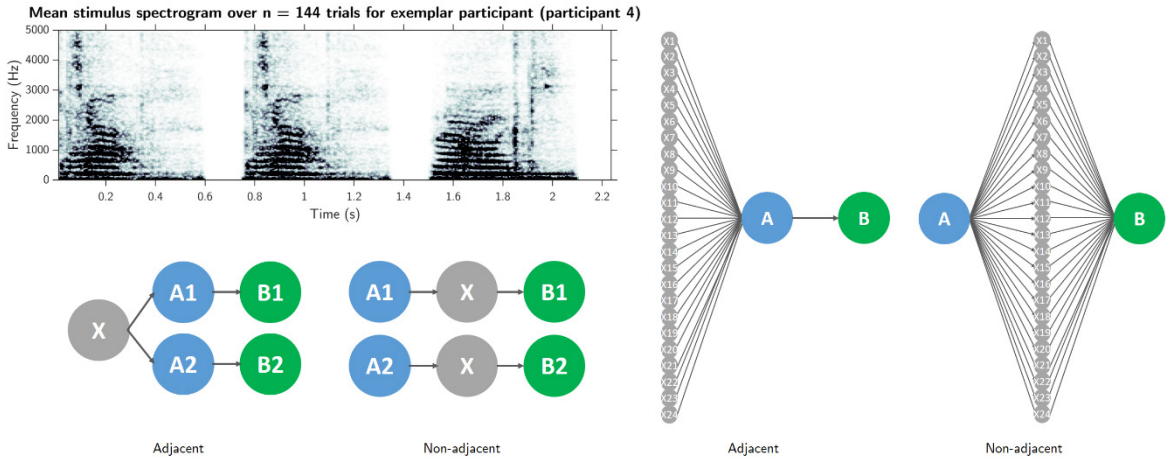


Figure 2.3 **The *AxB* task grammar and representative mean stimulus spectrogram.** The lower left schematics illustrate all valid sequences by abstracting over the category of irrelevant **X**s but showing the critical **A→B** relationships. On the right, the same grammar is illustrated with the **X**s expanded to better highlight the adjacent and non-adjacent dependencies (note, however, that **A** and **B** categories are now abstracted for ease of visualisation).

Having two valid **A→B** pairs in the grammar made it possible to generate invalid sequences using the same vocabulary as valid sequences whilst balancing word frequency and order. Additionally, and crucially in consideration of this study's role as a precursor to a prospective electrophysiological task, it also meant that, during testing, the grammaticality of the **B** element could

be manipulated whilst its acoustic features remained unchanged, since it could either conform to or violate a previously established **A→B** expectation (an approach followed in prior electrophysiological AGL studies, for example Kikuchi et al., 2017). Table 2.2 shows the final 2×2×2 factorial design and trial breakdown for each balanced stimulus condition.

Table 2.2: **Testing phase trial breakdown for the 2×2×2 factorial design of the deterministic AG task.** In this cohort, there were two test runs of 48 trials each. During the exposure phase, only consistent trials were presented (4 left-hand cells), but with 12 trials per category instead of 6.

Per test run (48 trials total)	**Consistent trials: 24**		**Violation trials: 24**	
Adjacent trials: 24	**xA_1B_1:** 6 trials	**xA_2B_2:** 6 trials	**xA_1B_2:** 6 trials	**xA_2B_1:** 6 trials
Non-adjacent trials: 24	**A_1xB_1:** 6 trials	**A_2xB_2:** 6 trials	**A_1xB_2:** 6 trials	**A_2xB_1:** 6 trials

Having described the artificial grammar and the main exposure and testing phases of the task, it is now possible to describe and motivate the passive *pre-exposure* phase that preceded them. The pre-exposure phase was designed as a precursor to the more typical sequence exposure phase, specifically to boost learning of the full grammar by first promoting the generalisation of the **AB** pair knowledge to novel sequences. Incrementally ordered training can enhance learning in artificial grammar tasks (Poletiek et al., 2018).

During the pre-exposure phase, the participant was passively exposed to 24 trials per valid **AB** pair (i.e. 24 x 2 = 48 trials). Each trial consisted of three short related sequences, each separated by 900ms of silence. Specifically, every trial followed the pattern

$$A_iB_i \ \langle\text{silence}\rangle \ x_jA_iB_i \ \langle\text{silence}\rangle \ A_ix_kB_i,$$

always in that order. Like the pre-exposure phase itself, this pattern follows the doctrine of "starting small" (Elman, 1993), whereby simple relationships are established before complex ones. As before, the duration of each syllable was 600ms and syllables were separated by 150ms of silence.

Thus, each three-element sequence was 2100ms long, each two-element sequence 1350ms long, and the pre-exposure phase around 6 minutes in total. As above, letters here denote a broad class of stimuli, whilst subscripts indicate a specific stimulus of that class. Thus, **A** and **B** did not differ across the three sequences, but **X** did. Furthermore, across all 24 trials generated per **AB** pair, every possible **X** was featured twice, once as $\mathbf{X}_j$ and once as $\mathbf{X}_k$.

2.3.4 Post-task questionnaire

Following completion of the task, participants were provided with a questionnaire (see Appendix 3: Behavioural questionnaire). This questionnaire incorporated questions on basic inclusion/exclusion criteria, demographic information, and potential confounds to task performance. It also included two questions designed to establish to what extent, if any, participants developed an explicit awareness of grammar rules or their own performance.

Finally, two additional questions were included to assess the impact of a potentially significant factor affecting future neurosurgical testing on this paradigm: subject wakefulness. It has been long-established that surgical inpatients are more sleep deprived than healthy controls, mainly by dint of the hospital environment (Cumming, 1984). Furthermore, a comprehensive post-implantation medication regimen can potentially impact further on wakefulness, either as a result of known side-effects of specific drugs, for example opioid analgesics or antiepileptics (Rogers et al., 2013; Salinsky et al., 1996), or by way of the measurable impact of polypharmacy on sleep quality and sleep stage durations (Lande & Gragnani, 2015). Finally, in some hospitals, sleep deprivation is an intentional component of the monitoring procedure for epilepsy patients since it has a potentially epileptogenic influence, although this is typically used as a tool to increase diagnostic yield in early testing rather than during invasive monitoring (Giorgi et al., 2013). Thus, considering the many potential sources of disruption to wakefulness in a future electrophysiological experiment, it was considered appropriate to assess any possible relationships between fatigue/sleep deprivation and performance on the *AxB* task.

2.3.5 Analysis

Behavioural task data were automatically recorded by the task script and subsequently analysed within MATLAB. Questionnaire responses were likewise transferred to MATLAB for statistical analysis.

Throughout the analysis, nonparametric tests such as the Wilcoxon signed rank test were used to compare group means in cases where data was found to be non-normally distributed in at least one group, as assessed by visual inspection and formal methods (rejection of null hypothesis of normality, $p < .05$, Lilliefors test). The exception is in the use of *n*-way factorial ANOVAs, for which no non-parametric equivalent is defined, and which have been shown to be robust to violations of the normality assumption (Blanca et al., 2017).

2.4 Results

2.4.1 Overall performance

The primary behavioural performance measure was the proportion of correct responses to the testing trials of the task (that is, the proportion of correct grammaticality judgements). The full cohort of 50 participants performed significantly better than chance at the group level, averaged across both testing runs ($\mu = 0.71$, $p < .001$; one-sample Wilcoxon signed rank), demonstrating that the cohort as a whole was able to distinguish grammatical from ungrammatical sequences.

A four-way mixed-design analysis of variance (ANOVA) was conducted to reveal differences in task performance associated with manipulations to one or more of the experimental factors (between-participant factor: Prosodic group; within-participant factors: Run, Adjacency, Grammaticality; interactions included). The ANOVA revealed no statistically significant effect of the prosodic group on performance ($F_{2,47} = 0.132$, $p = .877$, although this result is revisited in more detail in subsection 2.4.3). Because it was a key between-participant manipulation, and despite the lack of an observable prosodic effect, the distribution of scores under each condition is nevertheless shown in Figure 2.4. The figure demonstrates that there were no clear performance benefits for the participants exposed to prosodic cues, relative to those provided with no additional cues.

Likewise, the ANOVA revealed no significant effect of adjacency ($F_{1,47} = 0.377$, $p = .542$), and no significant interactions between any combination of factors ($p > .05$ in all cases). However, the ANOVA revealed significant effects of grammaticality ($F_{1,47} = 5.94$, $p = .019$) and run ($F_{1,47} = 25.8$, $p < .001$) on performance. Consequently, post-hoc pairwise analyses were conducted to clarify differences in performance under each grammaticality and run condition.

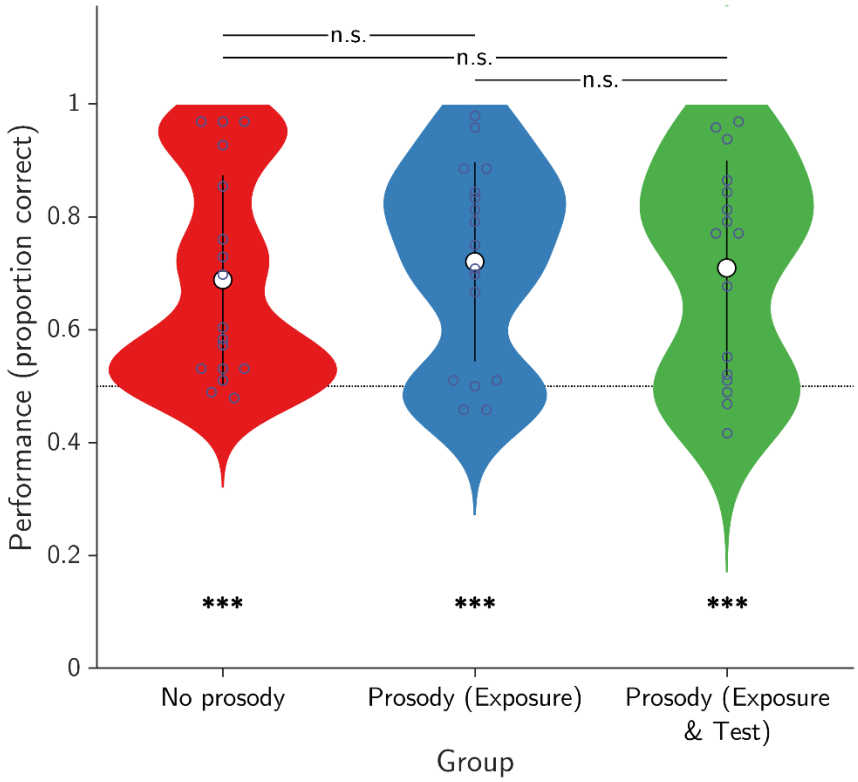


Figure 2.4 **Comparison of scores based on presence of prosodic pitch cues.** Each violin plot shows the distribution, mean and standard deviation of scores averaged across the two testing runs under each manipulation of prosodic pitch (NP: 'no prosody'; PE: 'prosody during exposure only'; PT: 'prosody during both exposure and testing'). Differences between the mean score across runs under the three conditions were non-significant ($\mu_{PE} = 0.72$, $\mu_{PT} = 0.71$, $\mu_{NP} = 0.69$). However, all groups scored significantly higher than chance overall. *** = $p < .001$.

A nonparametric comparison between mean performance across runs on grammatical versus ungrammatical sequences revealed significantly higher performance on grammatical sequences (Wilcoxon signed rank, $\mu_{Gram} = 0.74$, $\mu_{Ungram} = 0.67$, $p = .038$). A nonparametric test was used here because "ungrammatical" scores were non-normal according to a Lilliefors test ($p = .019$). Despite the difference in scores between the grammaticality conditions, performance on both grammatical and ungrammatical sequences was significantly above chance (one-sample Wilcoxon signed rank tests, $p < .001$ in both cases). Figure 2.5 illustrates this. Comparing the proportions of trials each participant responded "grammatical" or "ungrammatical" revealed a statistically significant difference between them at the group level (Wilcoxon signed rank, $p = .038$; $\mu_{responded\ gram} = 0.53$; $\mu_{responded\ ungram} = 0.47$), showing a small but significant response bias towards "grammatical". This response bias in at least some participants explains the tighter distribution and higher scores on "grammatical" sequences shown in Figure 2.5.

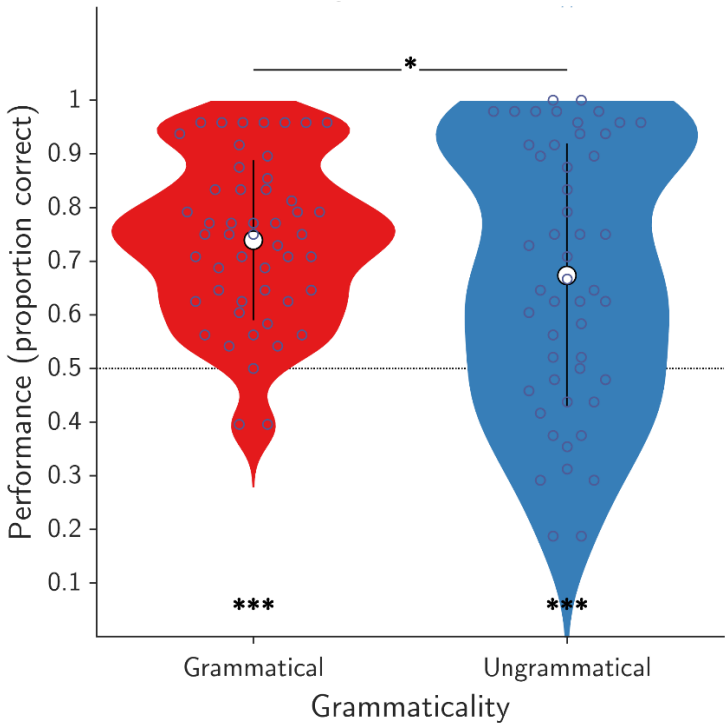


Figure 2.5: **Mean performance across task runs on grammatical versus ungrammatical sequences.** Violin plots show the distribution of performance and probability density for grammatical and ungrammatical trials, across both runs and the entire cohort (n = 50). Overlays show the mean and standard deviation. Performance was significantly higher for grammatical than ungrammatical sequences, but was nevertheless significantly higher than chance under both conditions. Significance computed using one-sample and two-sample paired Wilcoxon signed rank as appropriate. * = $p < .05$, *** = $p < .001$.

Performance was above chance on both testing runs ($p < .001$ in both cases; one-sample Wilcoxon signed rank tests), indicating that the grammar was learned by the group by the end of the full exposure period. However, a comparison between performance on both runs revealed that, across all participants, performance was significantly higher on Run 2 than Run 1 (Wilcoxon signed rank, $p < .001$, $\mu_1 = 0.65$, $\sigma_1 = 0.18$; $\mu_2 = 0.77$, $\sigma_2 = 0.22$; both groups non-normally distributed, i.e. $p < .05$, Lilliefors test). These results are illustrated in Figure 2.6.

Looking more closely at the data, on a per-participant basis, it transpired that 21 of 50 participants scored significantly higher than chance in the first run, as determined by two-tailed comparison with the inverse binomial cumulative distribution function at $\alpha = .05$. In the second run 29 participants scored significantly higher than chance, by the same method. Group performance in

each run of the task was thus driven by a subset of participants performing at a high level. By the second run, the division between these subsets produced a clearly bimodal distribution (see Figure 2.6).

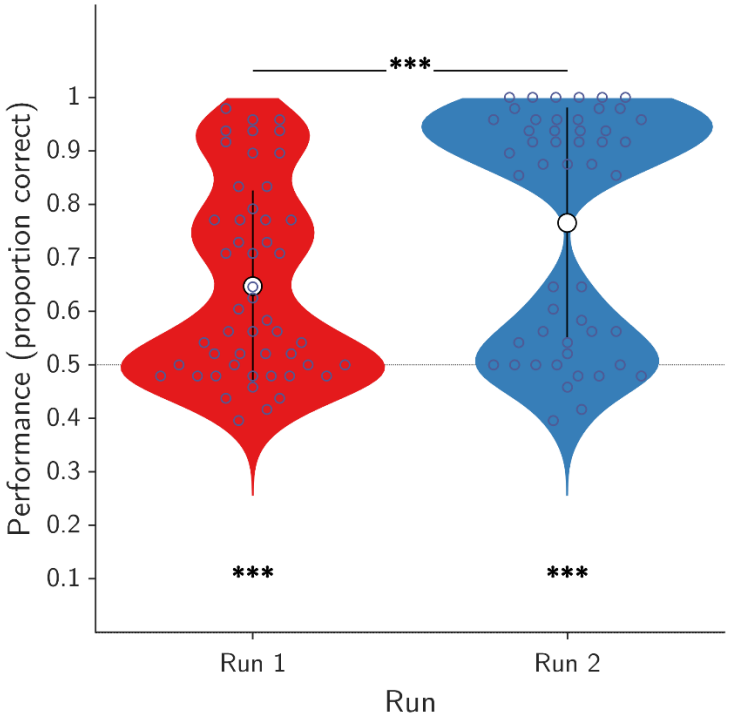


Figure 2.6: **Performance (proportion correct) on both runs of the *AxB* task.** Violin plots show the distribution of performance and probability density for the first and second runs of the task's testing phase, across the entire cohort (n = 50). Overlays show the mean and standard deviation for each run. Performance on both runs was significantly different from chance, but significantly improved between the first and second runs. Performance was clearly bimodal by the second run. Significance computed using one-sample and two-sample paired Wilcoxon signed rank as appropriate. *** = $p < .001$.

2.4.2 The bimodal distribution of performance

Given the bimodal distribution of the data in the final run of the task, performance was described by fitting a Gaussian mixture model to the distribution of final run scores using maximum likelihood estimation, yielding two normal distributions: $\mu_1 = 0.52$, $\sigma_1 = 0.06$ (hereafter "chance performers") and $\mu_2 = 0.94$, $\sigma_2 = 0.05$ (hereafter "high performers"; see Figure 2.7). Though both Gaussians were cleanly separated as can be seen from the figure, each was associated with its own normal probability density function, meaning that for each participant, formal probability of membership to one or another distribution could be evaluated from their final run score. In this way, participants were each unequivocally assigned to the "high performer" or "chance performer" group

depending on which probability was higher. This method was used because it could usefully be applied to more equivocal distributions if needed in future. On this basis, 29 of 50 participants were high performers, and 21 of 50 chance performers, which accords with the binomial significance testing on individual participant's scores reported above. The fact that 58% of participants reached close to ceiling performance, whilst 42% did not, is suggestive of the fact that the grammaticality judgement component of this task is non-trivial, in line with past findings (see 2.2.1). The binomial shape of the distribution suggests that, by the end of the task, participants either can, or cannot, operationalize their appreciation of the grammar, rather than awareness falling on a continuum.

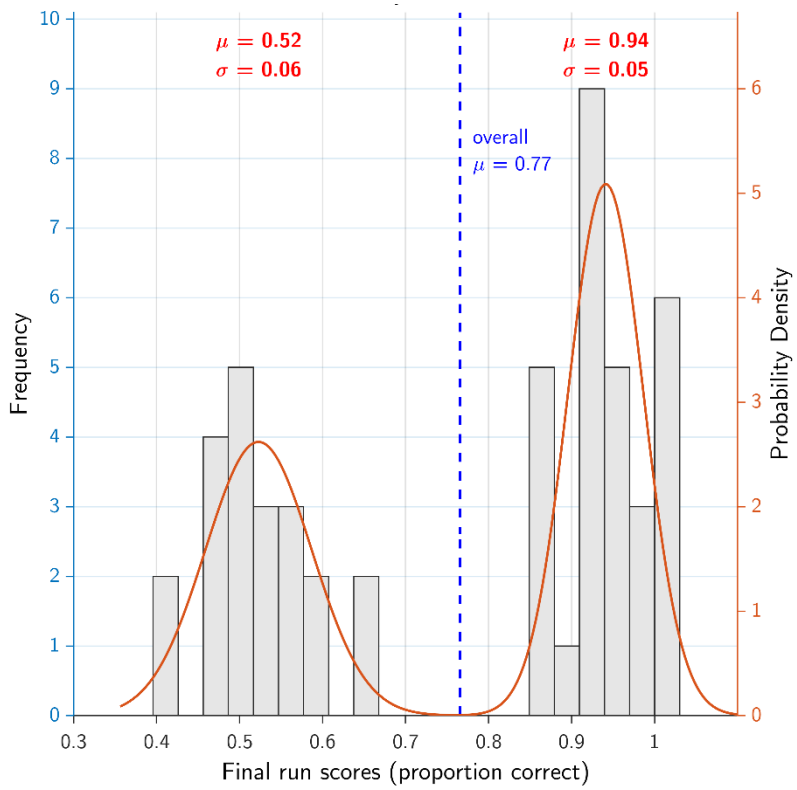


Figure 2.7: **Performance (proportion of correct responses) on the final run of the *AxB* task in healthy controls (N = 50)**. A histogram of performance across the deterministic cohort is shown overlaid with the normal distributions of a Gaussian mixture model fitted to the data by maximum likelihood estimation. The results exhibit a bimodal distribution with 29 "high performers", and 21 "chance performers", when assigning each participant to the distribution with the highest probability density for that participant's score.

2.4.3 Prosodic pitch effects in high and chance performers separately

The 4-way ANOVA reported previously revealed no significant effect of the prosody condition on performance ($F_{2,47} = 0.132$, $p = .877$), and no significant interactions between prosody and any other factors. However, as shown above, by the final run of the task, the cohort could be cleanly segregated into high and chance performers, which it was considered might reveal differential prosodic effects. Further analysis of prosody was therefore undertaken, focussing on the final run of the task alone.

Looking at the final run scores, all three prosodic groups performed significantly better than chance overall (one-sample Wilcoxon signed rank tests; No prosody: $p = .0014$, $\mu_{NP} = 0.73$, $\sigma_{NP} = 0.22$; Prosody in exposure: $p < .001$, $\mu_{PE} = 0.79$, $\sigma_{PE} = 0.21$; Prosody in testing and exposure: $p < .001$, $\mu_{PT} = 0.77$, $\sigma_{PT} = 0.22$). There were no statistically significant differences between the scores of these groups in the final run (Kruskal-Wallis test, $\chi^2 = 0.33$, $p = .85$).

As in the overall analysis, in each prosody condition, participants' scores followed a bimodal distribution (see Figure 2.8). Therefore, the condition comparison was additionally undertaken for the high performers and chance performers separately, to test whether prosodic manipulations were effective in boosting performance in one or another of these groups despite being non-significant overall. Despite this, one-way analysis of variance (ANOVA) revealed that, for both chance performers ($F_{2,18} = 0.30$, $p = .74$) and high performers ($F_{2,26} = 0.86$, $p = .43$), the effects of the manipulation remained non-significant. As a result, for the purposes of further analysis, all participants were grouped across these three prosodic conditions.

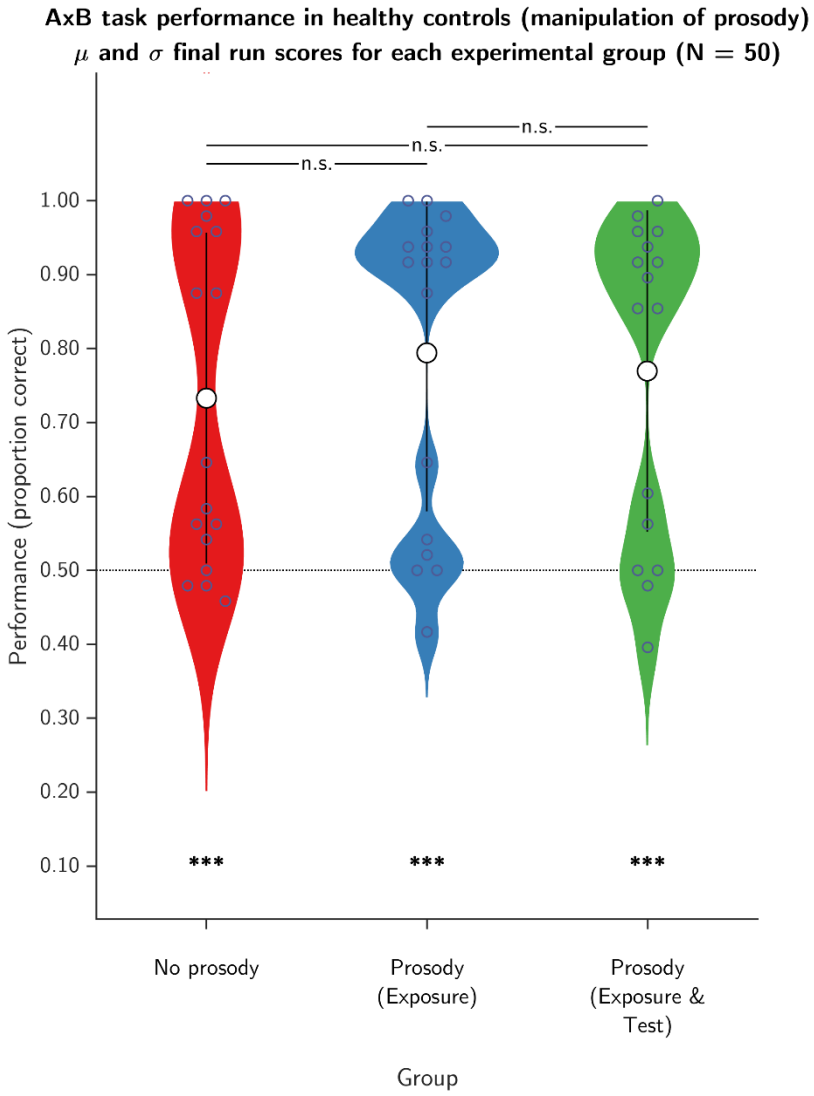


Figure 2.8 **Comparison of final run scores based on presence of prosodic pitch cues.** Each violin plot shows the distribution, mean and standard deviation of the scores under each manipulation of prosodic pitch (NP: 'no prosody'; PE: 'prosody during exposure only'; PT: 'prosody during both exposure and testing'). Whilst the means for the two prosodic manipulations (μ_{PE} = 0.79, μ_{PT} = 0.77) were marginally higher than those for the non-prosodic condition (μ_{NP} = 0.73), differences between them remained non-significant. However, all three groups scored significantly higher than chance overall. *** = $p < .001$.

2.4.4 Correlation of performance under opposing conditions

The reported 4-way ANOVA described group-level effects of the experimental conditions on performance, but could not reveal how scores varied under stimulus conditions on a per-participant basis. Whilst at the group level sequence adjacency had no effect on performance, for example, it might nevertheless be the case that individual participants exhibited a trade-off in performance between the opposing stimulus conditions. This could occur, for example, if there were competition for cognitive resources between the two sequence types. Scores under opposing stimulus conditions ("adjacent" versus "nonadjacent"; "grammatical" versus "ungrammatical") were therefore correlated to determine the strength and direction of any linear relationship between them.

The results, summarised below in Figure 2.9, showed a significant positive correlation between scores under key opposing stimulus conditions (adjacent vs. nonadjacent: $r = 0.89, p < .001$; grammatical vs. ungrammatical: $r = 0.69$, $p < .001$). Overall, this meant that relatively higher performing participants under one condition also tended to be higher performers under the opposing condition.

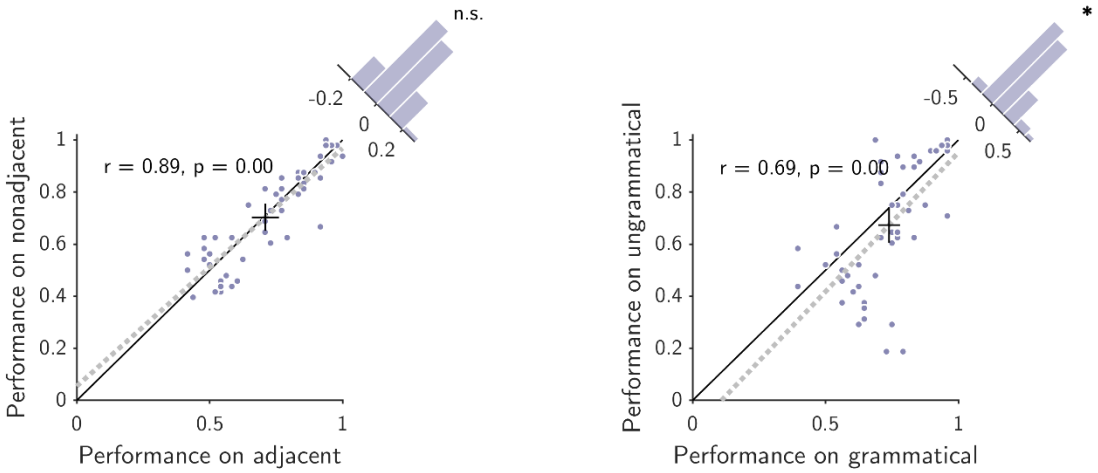


Figure 2.9: **Correlation and comparison of within-subject performance under opposing stimulus conditions.** Results are shown for the full cohort (n = 50). Each subfigure includes a scatterplot showing performance in every participant, along with the Spearman rank correlation coefficient and significance. The black "+" indicates the centroid of the data, scaled to reflect 95% confidence intervals along both axes, under a normality assumption. A least squares fit (dashed) is also shown along with a line of identity (solid black) for reference. Finally, a histogram shows the distribution of data relative to this line of identity, along with a summary of a Wilcoxon signed rank test between results under the two conditions. Both correlations were significantly positive. Differences between conditions were significant in the case of grammatical vs. ungrammatical (* = $p < .05$).

In the case of adjacent and nonadjacent sequences, these results clearly suggested there was no performance trade-off between the two sequence types. However, it was apparent from the distribution of the grammaticality scores that there were nonlinearities in the co-distribution of scores caused by aggregating both high and low performers (see scatter points, Figure 2.9, right-hand plot). The analyses above were therefore repeated after once again segregating the high and chance performers. These results are reported in the following section.

2.4.5 Correlations in high and chance performers separately

Taking only the high performers, and repeating the analysis in the previous section, the results obtained were similar to the cohort overall (see Figure 2.10), demonstrating significant positive Spearman rank correlations between the opposing stimulus conditions (adjacent and nonadjacent: $r = 0.79$, $p < .001$; grammatical and ungrammatical: $r = 0.52$, $p < .001$). There were no significant differences between performance under opposing conditions (paired t-tests, $\alpha = .05$). Here, as for the cohort as a whole, it is evident that high performers are not just generally good at the task, but that performance under opposing stimulus conditions is similar and positively correlated.

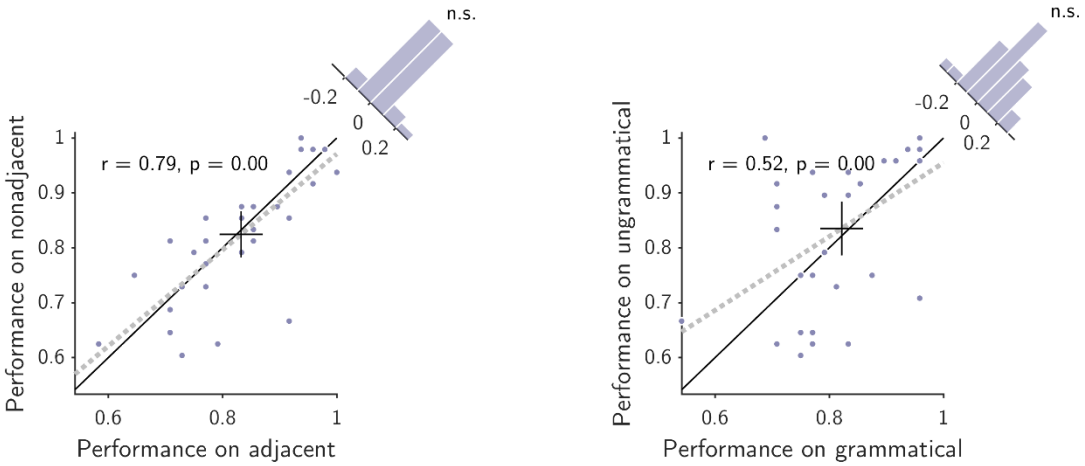


Figure 2.10: **Correlation and comparison of within-subject performance (high performers)**. There were n = 29 high performers. In this group, differences between scores under opposing conditions were non-significant (here, paired t-tests, $\alpha = .05$), and correlations between them significantly positive (see main text).

The picture was very different in chance performers (see Figure 2.11). Here, scores under opposing adjacency conditions were uncorrelated (r = -0.22, p = .36) and scores under opposing grammaticality conditions were significantly negatively correlated (r = -0.73, p < .001). There were also significant differences between performance on grammatical versus ungrammatical sequences (μ_{cons} = 0.60, μ_{vio} = 0.41, paired t-test, p < .001).

This general pattern greatly differs from the pattern shown by high performers. The uncorrelated adjacency distribution is suggestive of random guessing by the chance performers (left, Figure 2.11). The significant negative correlation in the grammaticality distribution (right, Figure 2.11) could be suggestive of a cognitive trade-off between performance under opposing conditions. However, the significant performance difference between grammaticality conditions means this likely reflects a bias towards judging sequences as "grammatical" superimposed on random guessing behaviour. Here, the distribution is constrained along the downward diagonal because it consists only of participants whose mean grammaticality judgement scores did not significantly differ from chance. Accordingly, further analysis of response frequencies revealed no

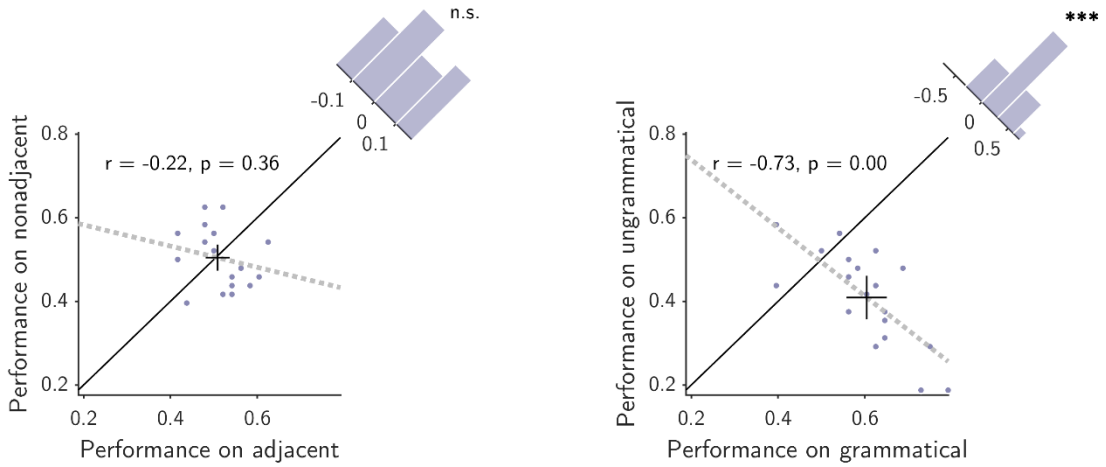


Figure 2.11: **Correlation and comparison of within-subject performance (chance performers)**. There were n = 21 chance performers. Despite overall chance performance, the range of scores on specific conditions was considerable, as the grammaticality distribution (right) shows. Differences between scores under opposing conditions (paired t-tests, α = .05) were significant for grammatical vs. ungrammatical (p < .001), but non-significant for adjacency (p = .85). This suggests a simple response bias in favour of "grammatical" judgements. Correlations between scores under opposing conditions were negative, shown by the dashed least-squares fit lines, although this was only significant in the case of consistent vs. violation (p < .001).

response bias in high performers, but a significant bias towards responding "grammatical" in chance performers (paired t-tests; high performers: $p = .67$, $\mu_{respondedgram} = 0.50$; chance performers: $p = .0012$, $\mu_{respondedgram} = 0.59$).

2.4.6 Explicitness of the task and impact of wakefulness

AGL paradigms were initially designed to measure implicit learning, and this is often assumed to be the case. However, recent approaches have challenged the "implicitness" of grammaticality judgement tasks, which require explicit decision making and conscious reflection. To attempt to assess awareness of their knowledge, we asked participants to complete a questionnaire about what they had learned. Questionnaire responses were coded and analysed to determine the relation of performance (proportion correct) to subjective confidence, and its relation to explicit knowledge of the grammar (number of rules identified). The impact of participant wakefulness was also assessed. Analysis was accomplished by computing Spearman rank correlation coefficients between cohort performance and all coded subjective questionnaire responses.

Significant correlations are summarised in Figure 2.12. These firstly reveal that the primary performance measure is correlated with subjective assessment of performance on the task ($r = 0.82, p < .001$), and with the number of rules of the artificial grammar the participant could recall and correctly specify ($r = 0.83, p < .001$). As such, under the *AxB* task, the grammaticality judgement performance is seemingly reflective of both awareness of subjective performance, and of the rules of the artificial grammar, both of which are forms of explicit knowledge.

Participant tiredness, as assessed by a subjective tiredness measure, was found to be significantly negatively correlated with performance on the task ($r = -0.32, p = .024$; see Figure 2.12). However, there was no significant correlation between performance and estimated number of hours slept ($r = 0.01$, $p = .95$), which probably simply reflects poor subjective awareness of restful sleep duration.

The significant relationship between subjective tiredness and performance is relevant to future deployment as an electrophysiological task in a potentially sleep-deprived patient cohort. Additionally, in conjunction with the results above, it is further suggestive of the fact that this AGL task, which notionally ought to be trivial for humans, is nevertheless remarkably cognitively demanding despite the inclusion of a number of aids to learning (see 2.2.2 and 2.3.2).

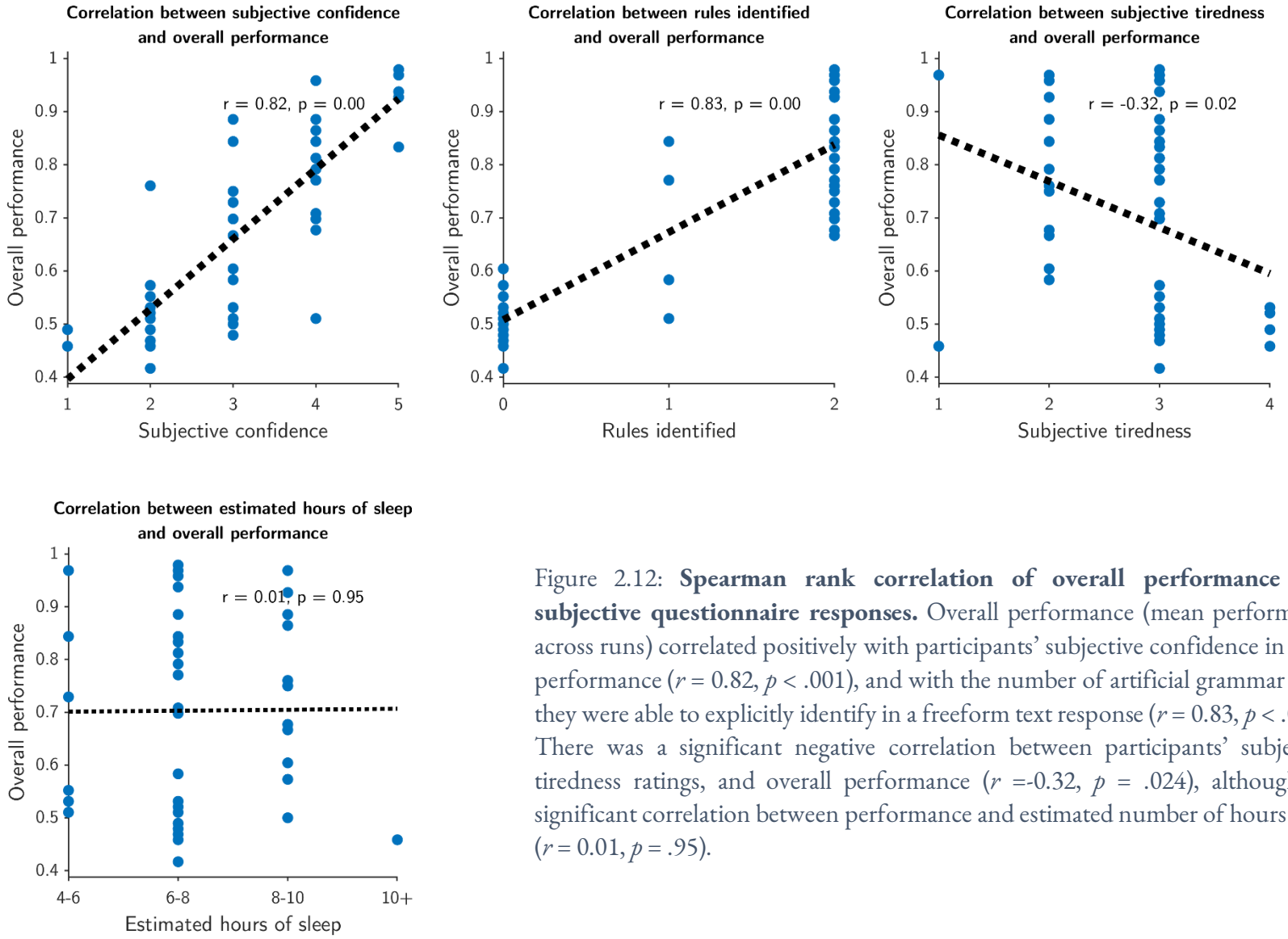


Figure 2.12: **Spearman rank correlation of overall performance and subjective questionnaire responses.** Overall performance (mean performance across runs) correlated positively with participants' subjective confidence in their performance ($r = 0.82$, $p < .001$), and with the number of artificial grammar rules they were able to explicitly identify in a freeform text response ($r = 0.83$, $p < .001$). There was a significant negative correlation between participants' subjective tiredness ratings, and overall performance ($r = -0.32$, $p = .024$), although no significant correlation between performance and estimated number of hours slept ($r = 0.01$, $p = .95$).

2.4.7 Reaction time differences

Reaction times (RTs) have the potential to reveal differences in cognitive processing that may not be apparent using other metrics such as performance. Prior analysis (see previous section) suggested that grammaticality judgement might be reflective of explicit awareness of sequence structure, yet the literature makes a compelling argument for the implicitness of statistical learning. Consequently, an analysis of RTs was undertaken with a view to determining firstly whether there existed evidence for reaction time differences related to the grammar, and secondly whether this was preserved even in participants unable to meet the explicit task demands. This being the case, it would suggest that RTs characterise an implicit sensitivity to the grammar under this task. RTs recorded during the task described, for each trial, the latency of a subject's button press following the jittered presentation of the visual cue (signalling onset of the response window) after sequence playback.

Mean RTs were computed and log-transformed for each condition in every participant, and compared using paired t-tests for the two manipulations of sequence structure (adjacency and grammaticality). For the cohort as a whole, the results were compelling (see Figure 2.13). Whilst there were no significant differences between mean RTs for the adjacency manipulation (paired t-test on log-transformed data, μ_{adj} = 1262 ms, μ_{nonadj} = 1245 ms, p = .31), the response latency to violation (ungrammatical) sequences was significantly longer, by an average of 129 ms, than that to consistent (grammatical) sequences (μ_{cons} = 1189 ms, μ_{vio} = 1318 ms, $p < .001$). Thus, it appears that reaction times can be related to sequence structure under the *AxB* task.

The same analysis was then conducted again for the high performers and chance performers separately, as previously. The results are shown in Figure 2.14. Notably, these show that in both high performers and chance performers, reaction times are systematically longer to violation sequences than to consistent sequences (high performers: $\mu_{vio} - \mu_{cons}$ = 128 ms, p = .0040; chance performers: $\mu_{vio} - \mu_{cons}$ = 129 ms, p = .034). The existence of RT differences for both high and chance performers, and the extremely similar magnitudes and directions of those differences in these two different subgroups of participants, constitute evidence that, in this cohort, reaction times reveal an implicit sensitivity to violations of the *AxB* grammar, even in participants unable to meet the apparently explicit demands of the grammaticality judgement component.

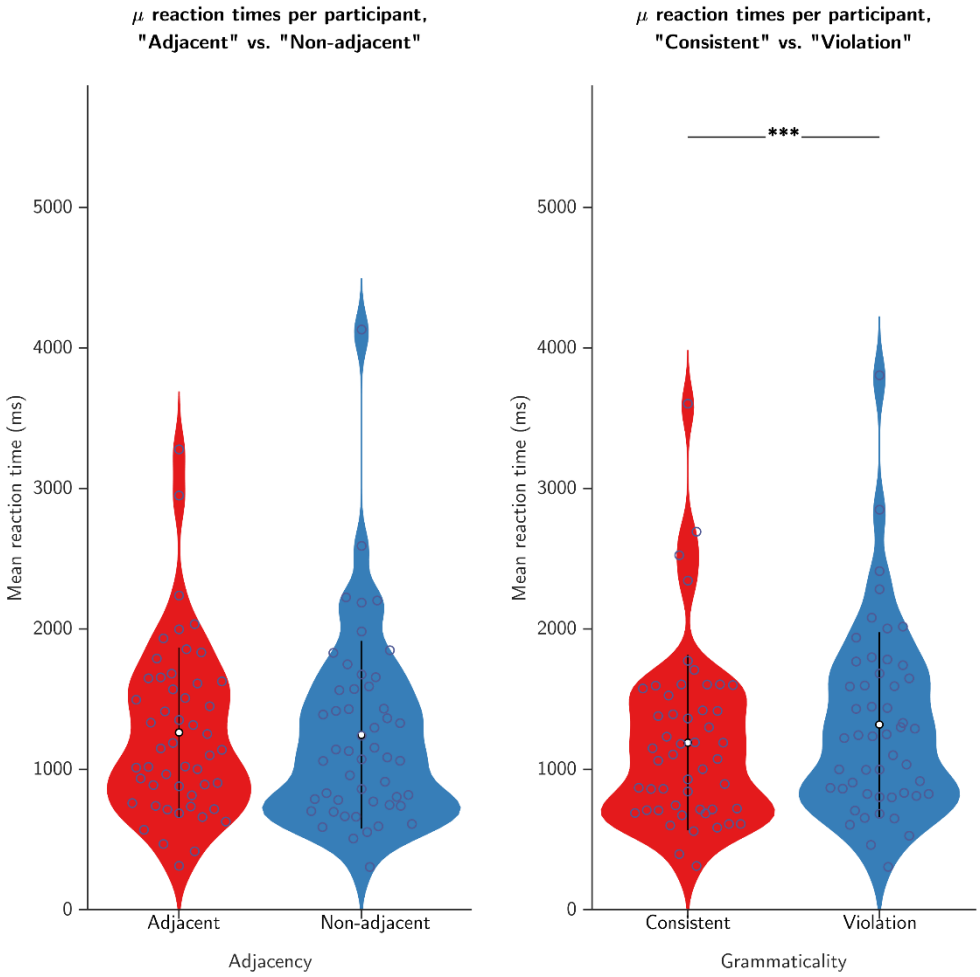


Figure 2.13: **Mean reaction times by structural condition (n = 50).** Shown are violin plots of reaction times (RTs) under the two manipulations of sequence structure. RTs were averaged across all trials of both testing runs in each participant. Each violin plot shows the individual data points, the probability density for each distribution, and is overlaid with the group mean and standard deviation. Group-level comparison of RTs revealed that responses to "violation" sequences were significantly delayed compared to responses to "consistent" sequences (paired t-test on log-transformed data, *** = $p < .001$, μ_{cons} = 1189 ms, μ_{vio} = 1318 ms).

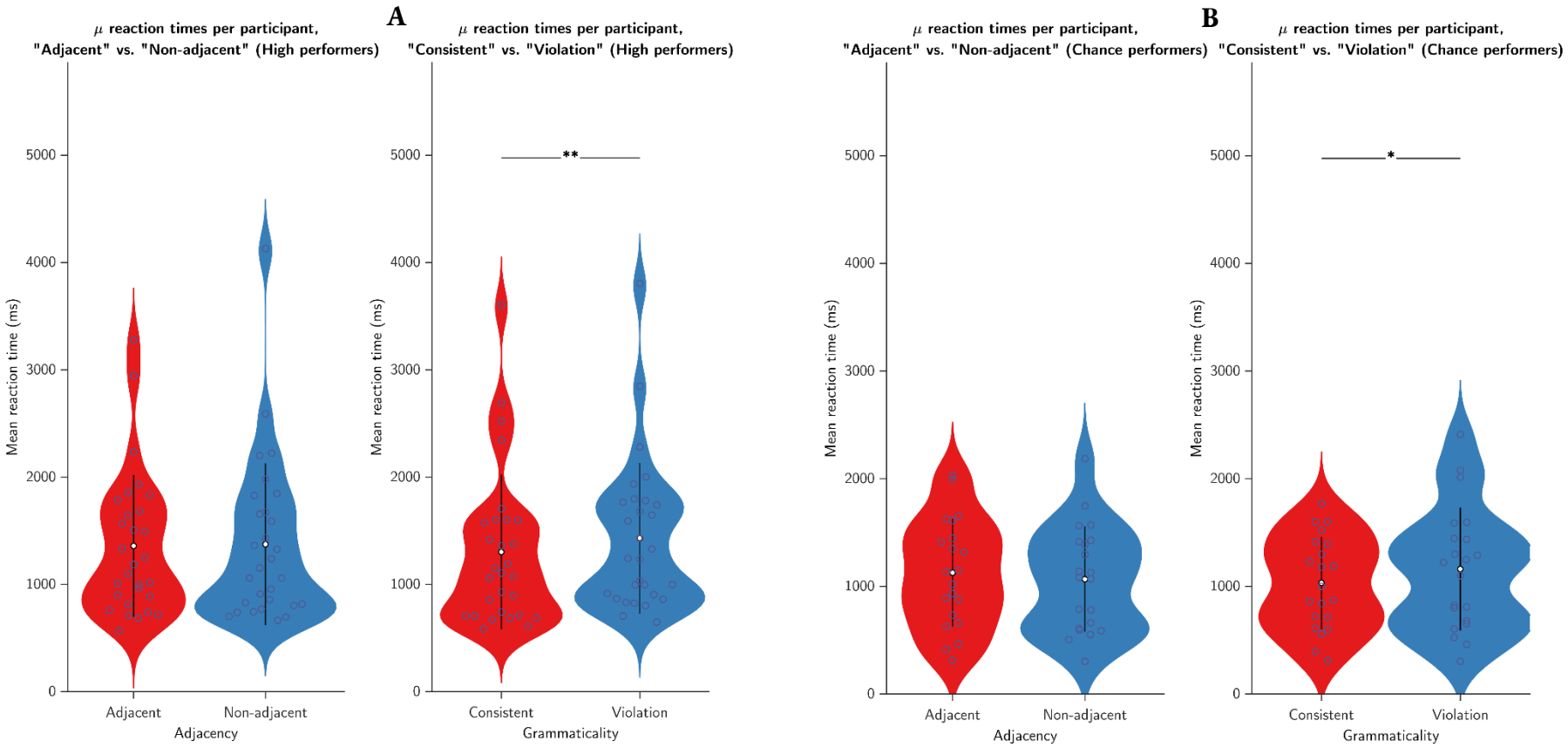


Figure 2.14: **Mean reaction times (RTs) for high performers (n = 29) and chance performers (n = 21).** As expected of a group whose scores demonstrated clear sensitivity to sequence grammaticality, high performers (panel A) exhibited the same reaction time effect as the group as a whole, having significantly longer RTs to violation vs. consistent sequences (paired t-test on log-transformed data, $p = .0040$, $\mu_{cons} = 1303$ ms, $\mu_{vio} = 1431$ ms). However, this effect was also observed in the chance performing group (paired t-test on log-transformed data, $p = .034$, $\mu_{cons} = 1033$ ms, $\mu_{vio} = 1162$ ms), suggesting implicit sensitivity to sequence structure under this task even for subjects who did not excel at making explicit grammaticality judgements. (* = $p < .05$, ** = $p < .01$)

2.5 Discussion

Despite their ubiquity in language, the ability to learn non-adjacent dependencies (NADs) appears to be highly selective compared to adjacent dependencies (ADs). The aim of the present study was to assess the ability of healthy adult participants to learn regularities in sequences generated from a novel artificial grammar containing both adjacent and non-adjacent dependencies. This approach allowed us to conduct within-participant comparisons of performance on different types of sequences.

2.5.1 Overall performance

Our results revealed that participants were able to learn the artificial grammar, performing well above chance on the grammaticality judgement task at the group level, across all trials. Moreover, this learning occurred in spite of the absence of segmental phonological cues to the roles of words in the artificial grammar, a key cue present in many existing studies. Performance was already above chance on the first testing run, implying that learning likely occurred as a result of the preceding period of exposure to grammatical sequences. The learning of relatively simple rules is expected following mere minutes of passive exposure (as shown, for example, by Saffran, Aslin, et al., 1996).

As revealed by an analysis of variance including four factors (test run, adjacency, grammaticality, prosodic group), significant effects were observed with respect to test run and grammaticality only (note that the lack of prosodic effect is discussed in more detail in subsection 2.5.2, followed by more detailed discussion of the effects of sequence structure, including adjacency). The significant effect of "test run" is expected; performance on the task improved significantly between runs, likely as a result of increased awareness of task demands over time, and the effectiveness of the intervening refamiliarisation period. However, surprisingly, grammaticality judgement performance was also shown to be significantly higher on grammatical sequences versus ungrammatical sequences. Interestingly, in a miniature language ("Mini Pinyin") learning study by Cross et al. (2020), which incorporated semantics, participants were similarly found to be more likely to respond correctly to grammatical versus ungrammatical sentences. However, the results show that the effect in our task was apparently caused by a response bias by the lowest performing participants, rather than a heightened ability to detect grammatical sequences.

Notably, despite the fact that the group was able to perform above chance on the task overall, performance followed a bimodal distribution; by the final run, around 60% of participants approached ceiling performance, but the remainder continued to perform around chance. Group performance was therefore driven by a subset of high performers, suggesting that the process of judging grammaticality was cognitively non-trivial. This discovery also prompted further analysis of the differential behaviour of chance and high performers that informs later discussion.

In a previous study on the concurrent learning of ADs and NADs, Romberg & Saffran (2013) found that familiarity judgement performance on NADs followed a bimodal distribution similarly centred around chance and ceiling levels, as ours did, whilst AD scores followed a unimodal distribution centred on middling scores. It is important to note that, under their task, NADs and ADs were formed between different dependents, rather than between the same dependents optionally separated by an intervening item, as under our task (see 2.5.3). However, in the grammar Romberg and Saffran used, the assessed NADs consisted of only 3 deterministic word pairs, whilst the ADs consisted of many more probabilistic pairings, making the NADs the more reliable patterns. Whilst the results were interpreted by the study authors as showing that maximum performance on the reliable NADs exceeded that on the unreliable ADs, it might rather be said that it is remarkable, as under our task, that such strongly highlighted dependencies were not simply associated with unimodal ceiling performance.

2.5.2 The effect of prosody on dependency learning

Prosodic pitch appeared to have no significant effect on performance under the task in the cohort overall, or in chance- and high-performing participants separately. There are a number of potential reasons for this. Firstly, it is possible that a net 10% pitch difference between important (**A**/**B**) and unimportant (**X**) items was too insubstantial to perceptually highlight the dependency in the context of the task. However, this manipulation exceeded previously reported psychophysical thresholds (Liu, 2013). Furthermore, equivalent proportional pitch differences, applied as here to naturalistic speech frequencies, have been shown previously to elicit electrophysiological responses in adults and infants within a deviance-detection task (Mueller et al., 2012), and likewise shown to serve as an assistive cue to infants undertaking an artificial grammar learning paradigm (Gervain & Werker, 2013). It therefore seems unlikely that the manipulation was insufficient as a cue in the context of this task.

It is also possible that prosodic saliency was not a significant aid to learning because of the other cues present within the task. The ~40% chance-performing participants might have been hampered not by the process of isolating salient items, but by a difficulty recalling one or more of the correct word pairings after exposure, a failure to appreciate the task requirements, or some other limitation. Likewise, in the subset of participants approaching ceiling levels on our task, it is possible that staged pre-exposure (discussed in Elman, 1993) and variable intervening items (Gómez, 2002) were sufficient to highlight the important elements and that prosodic pitch could add little to the saliency. Silent pauses within and around the sequences likely produced edge effects (as reported in Peña et al., 2002, for example, and discussed in Wilson et al., 2018) that acted as cues to saliency, for example, likely diminishing the relative benefits of prosodic pitch or other cues. However, it is important to note that prosodic pitch had a discernable effect in both Mueller et al. (2012) and Gervain & Werker (2013), even though the former used separate syllables surrounded by pauses, as in this task, and the latter employed a continuous speech stream.

Finally, it is conceivable that the prosodic comparison was conducted between groups too small to detect a significant effect. Further characterisation of pitch effects under a task solely focussed on prosody might be warranted in future. However, bearing in mind this study's goal of future electrophysiological deployment, the prosodic manipulation lacks utility. Groups here contained at least 16 participants; in an intracranial study, neurosurgical cohort sizes are restricted by clinical availability and are typically smaller than this. The manipulation is therefore unlikely to enhance learning in an intracranial study based on this task.

2.5.3 Performance by stimulus condition

Under our mixed-dependency grammar, group performance did not differ between adjacent and non-adjacent trials, and moreover individual scores on both types of sequence were correlated. This is perhaps unexpected in light of existing studies that show that AD learning is far less selective than NAD learning (Gebhart et al., 2009) and that NAD performance strongly relies on the presence of perceptual cues (Newport & Aslin, 2004; Onnis et al., 2005). However, unlike the cited studies, our task incorporated staged learning (Poletiek et al., 2018; Lai & Poletiek, 2011; Elman, 1993) and distributional cues (in particular high intervening element variability; Gómez, 2002), which are known to assist NAD learning.

Furthermore, very few artificial grammar learning studies have overtly tested concurrent adjacent and non-adjacent dependency learning in humans (Deocampo et al., 2019; Vuong et al., 2016; Romberg & Saffran, 2013). In these studies, results varied. Romberg & Saffran (2013) found that there was a trade-off between the simultaneous learning of ADs and NADs by individuals under conditions of high item variability, when ADs were probabilistic and NADs were nondeterministic, whilst online testing by Vuong et al. (2016) under a serial recall task revealed no significant differences between NAD and AD sensitivity under conditions of low variability. Deocampo et al. (2019) found that AD learning was more robust than NAD learning, as did Vuong et al. (2016) during offline testing, whilst Romberg & Saffran (2013) found that NADs were more salient. Critically, however, none of these studies incorporated ADs and NADs between *identical* dependent items (i.e. **xAB** or **AxB**, a situation that also occurs in language; Deocampo et al., 2019), randomly intermixed across trials. It is possible that, by incorporating shared dependents across multiple distances, the adjacent and non-adjacent relationships can be generalised to produce a single cognitive representation of a flexible relationship over a variable distance (**A→B**). Likewise, the ecologically valid "starting small" approach probably serves a similar purpose, highlighting a non-adjacent dependency by anchoring it to an earlier adjacent representation (Lany et al., 2007; Lany & Gómez, 2008; Lai & Poletiek, 2011). Both design constraints might explain why, under this task, NAD and AD performance is matched and correlated.

Previous statistical learning studies have suggested that subjects are extremely sensitive to the probability of temporally transitioning from an item to the one immediately following it (*first order* transitions), with even infants apparently able to use these statistics to group items into word-like units (Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996; Aslin et al., 1998). Humans have also been shown to be sensitive to higher order statistics, potentially up to third-order statistics (those that take into account a history of three preceding items; discussed in Cleeremans & McClelland, 1991). However, by generalising learning across distances, a subject may avoid needlessly integrating higher-order transitional information to process a non-adjacent dependency, but rather simplify the task by filtering out irrelevant or unreliable information, such as low-probability intervening elements, and assessing what remains using a simple first-order system. This would be more computationally efficient and effective than attempting to integrate probabilistic transitional information over multiple items (i.e. attempting to resolve **A→x** and **x→B** or even $p(\mathbf{B} \mid \mathbf{A}, \mathbf{x})$ to

assess the likelihood of A→x→B), an approach which could not easily support detection of relationships spanning low- or zero-probability intervening words, such as those featured in this task (and in real sentences).

2.5.4 Implicitness of the task and its measures

Implicit learning is the process of learning without awareness. Historically, AGL tasks were conceived as a means of studying implicit learning, but the implicitness of these tasks is the subject of ongoing discussion (DeKeyser, 2003). Here, grammaticality judgement accuracy was clearly correlated with questionnaire measures of explicit task awareness (both subjective confidence in the participant's own performance, and the number of grammatical rules consciously identified). This is strongly suggestive that the grammaticality judgement task draws on explicit awareness of the grammar, consistent with existing study suggesting it engages reflection-based decision processes (Christiansen, 2019). Eye tracking evidence (measuring regressions – right-to-left eye movements – during English reading by native and second-language speakers) likewise suggests that untimed grammaticality judgement tasks measure explicit task knowledge (Godfroid et al., 2015).

Despite this finding, the explicitness of a task metric does not imply explicitness of the entire task, which might invoke both explicit and implicit knowledge. Previous studies have demonstrated that indirect measures of sequence recognition such as reaction times may be more sensitive indices of learning than accuracy on explicit judgements (Batterink et al., 2015), and thus reveal implicit learning more effectively. This is a perspective strongly supported by our results, where reaction time effects were observed at the group level, but also separately within both the group who performed at ceiling levels on the explicit grammaticality judgement task, and the group of chance performers who did not demonstrate explicit knowledge of the grammar. Reaction times thus proved to be a sensitive measure of learning in this cohort, even in the absence of explicit awareness. Even the magnitude and direction of reaction time effects was remarkably similar in both the ceiling and chance performers, showing that the latency of responses to ungrammatical sequences was approximately 130 ms longer than that to grammatical sequences. This is consistent with an AGL study by Cock (2010), which found that response times were faster (specifically ~140 ms faster) for grammatical than ungrammatical test exemplars. However, it is generally difficult to meaningfully compare such effects across studies, given the very dissimilar demands placed on participants.

2.5.5 Suitability as an electrophysiological task

Analysis of the questionnaire responses showed that explicit task performance was significantly negatively correlated with subjective tiredness, which suggests that sleep-deprived neurosurgical patients (Cumming, 1984) will experience reduced grammaticality judgement performance relative to this cohort. However, as we have seen, most participants were able to concurrently learn the ADs and NADs within the grammar, as measured by grammaticality judgement performance, and even in those who only demonstrated chance performance, reaction time effects showed sensitivity to the grammar. In principle, therefore, this task has potential as an electrophysiological paradigm to investigate the neural substrates of structured sequence learning.

2.6 Conclusion

We found that participants could concurrently learn both adjacent and non-adjacent dependencies under a novel artificial grammar learning paradigm, exhibiting learning at a group level, and with more than half of participants performing at ceiling levels on a grammaticality judgement task in the presence of naturalistic cues to learning. Moreover, participants performed equally well on both dependencies, potentially generalising across them. We also showed that grammaticality judgement is both cognitively demanding and reflects explicit knowledge. Finally, we demonstrated that even in those participants performing at chance using the grammaticality judgement measure, they still exhibited reaction time effects of the same direction and magnitude as those seen in the ceiling performers, suggesting that reaction times reflect implicit learning of the grammar, and that implicit learning can still be observed in the absence of explicit knowledge of this grammar.

This study adds to the hitherto extremely limited literature on concurrent dependency learning, shedding light on the multifactorial influences affecting adjacent and non-adjacent dependency learning under a common task. Similarly, it adds to the limited literature assessing reaction time and grammaticality judgement measures as measures of implicit learning under a common task. Despite these findings, however, our understanding of the *neural mechanisms* supporting structured sequence processing remains limited. In light of the learning achieved here, it is therefore of vital importance that we better characterise human neural responses to this and other AGL tasks. This is the subject of the following chapter.

Chapter 3. Electrophysiological signatures of grammaticality in the human brain

The work presented in this chapter was primarily conducted by the author, except where explicitly stated. The author designed and implemented the sequence learning task in MATLAB following discussion with academic supervisors. The medical and surgical team at Iowa University Hospitals and Clinics undertook clinically indicated assessment, imaging and implantation of intracranial electrodes in human neurosurgical patients. Members of Iowa University's Human Brain Research Laboratory (HBRL) confirmed contact coordinates and regions of interest through co-registration and parcellation of structural imaging. The task was executed by our on-site colleague Zsuzsanna Kocsis, to whom we remain indebted, or by the author and HBRL staff members during collaborative visits. Subsequent electrophysiological analysis and writing were conducted by the author. Academic supervisors provided additional advice, comments and discussion.

3.1 Abstract

The brain is capable of learning covert structure over time without awareness. To measure these abilities, Artificial Grammar Learning (AGL) tasks expose participants to sequences conforming to a covert set of ordering rules (a grammar), and subsequently assess participants on their ability to distinguish novel "grammatical" sequences from "ungrammatical" sequences, for example by way of a grammaticality judgement task. These tasks can reveal sensitivity to proto-syntactic structure in the relative absence of semantic confounds. We previously revealed implicit and explicit sensitivity to grammatical violations in a healthy adult cohort under an auditory AGL task. The *AxB* task includes sequences containing dependencies between neighbouring items (adjacent dependencies) and between items separated by an uninformative intervening item (non-adjacent dependencies).

Key regions of a frontotemporal language network appear to exhibit distinct patterns of engagement in response to different types of dependency in both natural language and AGL tasks. In this study, we recorded behavioural and intracranial responses in 12 neurosurgical patients as they undertook the *AxB* task, in order to characterise neural responsiveness to sequence grammaticality manipulations at a high spatiotemporal resolution. Behavioural results did not provide confirmation of implicit or explicit sensitivity to grammaticality, likely due to confounds specific to the hospitalised patient population. However, previous research has revealed automatic neural responsiveness to syntactic manipulations even in the absence of conscious awareness. Consistent with this, time- and time-frequency domain analyses revealed statistically significant differences between neural responses to specific grammaticality conditions in regions including left inferior frontal gyrus (IFG) pars opercularis, right middle temporal gyrus and the right hippocampus.

Critical linguistic and AGL sequencing operations have been associated with canonical deflections of electroencephalographic (EEG) waveforms (ERPs), including two components associated with the processing of syntax, the left anterior negativity (LAN) and syntactic positive shift (P600). Here, we report a significant LAN-like response in the time-domain data within IFG pars opercularis, which has previously been implicated in the generation of these components, and more generally in the resolution of syntactic dependencies.

Overall, these results provide novel insights into the automaticity of grammaticality responses, and the location and syntactic relevance of classical ERP components under a non-linguistic task.

3.2 Introduction

The human brain exhibits an impressive facility for sequence learning, readily detecting statistical relationships between items presented over time in both the auditory and visual modalities (Henin et al., 2021; Milne et al., 2018; Stobbe et al., 2012). Remarkably, results collected under statistical sequence learning tasks have shown that we can learn to discern these statistical regularities incidentally, and seemingly without awareness, from mere exposure (Reber, 1967; Folia & Petersson, 2014), and even in a matter of minutes (Saffran, Aslin, et al., 1996; Saffran et al., 1997). Structural relationships (or *dependencies*) between ordered items also form part of the fabric of our everyday experience. These include dependencies between neighbouring items, and between items separated from one another (*adjacent* and *nonadjacent* dependencies). In particular, language incorporates complex dependencies, including adjacent and nonadjacent dependencies, at multiple structural levels. These dependencies obey specific rules (*syntax*) that ultimately become innately familiar to us, but which, much as in statistical learning tasks, we largely learn incidentally, and seemingly without awareness, from exposure to the utterances of others during development (Peter & Rowland, 2019).

A major source of evidence in the study of statistical sequence learning has been the artificial grammar learning (AGL) paradigm (Reber, 1967; Miller, 1958). During an AGL task, a subject is exposed to sequences of nonsense items ordered so as to conform to a covert grammar, and subsequently assessed on their ability to distinguish novel "grammatical" from novel "ungrammatical" sequences. Although originally intended specifically for the investigation of implicit learning (the process of learning without awareness), AGL tasks have been heavily utilised to study the learning of specific types of dependencies in sequences, in both humans and other species (ten Cate et al., 2020). Furthermore, these types of task continue to be vital to our efforts to characterise the neuronal substrates of sequence processing. Natural language exhibits a salience and structural complexity arguably unmatched by other sequential stimuli, but it contains more than syntax alone; it also contains meaning (semantics) and intent (pragmatics), and as such can evoke strong imagery or emotional responses that have the potential to recruit networks across "more or less the entire brain" (Hertrich et al., 2020). By contrast, AGL tasks can reveal responses to temporal structure in relative isolation from semantic influences (Petersson et al., 2012). The characterisation of human sequence processing abilities, and their neuronal underpinnings, thus benefits from the triangulation of evidence obtained under both statistical learning and natural language tasks.

Findings under natural language and statistical sequence learning tasks suggest the involvement of not only domain-specific mechanisms, but also a domain-general facility for proto-syntactic processing (Milne et al., 2018; Frost et al., 2015; Petkov & Jarvis, 2012; Conway & Christiansen, 2006). For example, despite much behavioural investigation into the extent to which sequencing knowledge can transfer between sensory modalities, there remains little consensus on the topic; some studies have provided evidence that cross-modal transfer does not readily occur (Conway & Christiansen, 2006; Frost et al., 2015; Li et al., 2018), and others have shown that it does (Kemény & Németh, 2017; Durrant et al., 2016). Whilst these results may not argue strongly for a wholly domain-specific or domain-general perspective, they hint at the potential relevance of both types of system. In line with this, other behavioural studies have provided clearer evidence for the involvement of a domain-general sequence learning system. Humans and other species exhibit commonalities in behaviour under sequence learning tasks across modalities (Seitz et al., 2007; Milne et al., 2018). Furthermore, performance on language tasks is known to be correlated with performance on more abstract sequence learning tasks (Misyak & Christiansen, 2012; Misyak et al., 2010; Conway et al., 2010; reviewed in Ettlinger et al., 2016). Finally, both developmental and acquired conditions have been shown to impact on both language and sequence learning abilities, despite relative sparing of other faculties. This has been demonstrated in participants with diagnoses including developmental dyslexia (Pothos & Kirk, 2004), specific language impairment (Evans et al., 2009; Hsu et al., 2014) and aphasia (Goschke et al., 2001; Christiansen et al., 2010; Cope et al., 2017). A body of behavioural evidence therefore suggests that domain-general processes contribute to sequence learning.

Neuroimaging evidence also supports the assertion that sequence learning and language are at least partly supported by domain-general operations. By comparing neural responses to natural sentences of varying syntactic complexity, imaging studies have revealed differential engagement of regions within a core system known to be involved in the processing of natural language syntax in humans, a left-lateralised *frontotemporal language network* (Friederici, 2011). This network appears to respond in a corresponding manner to manipulations of artificial grammar complexity (Friederici, Fiebach, et al., 2006; Friederici, 2011; Bahlmann et al., 2009; Petersson et al., 2012), suggesting that it has a broad role in the discernment of temporal structure. Thus, a mechanistic understanding of

this network appears relevant not only to our understanding of language, but also to cognition in general.

The frontotemporal language network comprises key areas in ventrolateral prefrontal, temporal and temporoparietal cortices (Friederici, 2011), including subregions of left inferior frontal gyrus (Brodmann Areas 44/45, incorporating Broca's area; Broca, 1861); the frontal operculum; superior temporal gyrus; and areas of the temporoparietal junction (TPJ), including angular gyrus and supramarginal gyrus (which correspond to parts of Wernicke's area; Binder, 2015; Wernicke, 1874, 1881). The inferior frontal gyrus (IFG), in particular, has long been associated with both language production and comprehension (Sahin et al., 2009; Broca, 1861), although its precise function is the subject of debate (Friederici, 2011). Its role in language has been variously surmised to encapsulate: perceptual conflict resolution (Hsu et al., 2017); general action observation and execution (Molenberghs et al., 2012; Pulvermüller & Fadiga, 2010; Rizzolatti & Arbib, 1998); verbal working memory, especially working memory for syntactic content (Yan et al., 2021; Rogalsky et al., 2008; Smith & Jonides, 1999; Caplan & Waters, 1999); and specific processes of syntactic unification and structure-building (Matchin & Hickok, 2020; Hagoort, 2005; Friederici, 2002; Grodzinsky, 2000). In particular, the engagement of ventral frontal regions, including IFG and frontal operculum, has been observed to vary with the complexity of sequencing operations, implicating frontal operculum in the processing of adjacent dependencies, and subregions of IFG in the processing of more complex dependencies (Friederici, Bahlmann, et al., 2006; Bahlmann et al., 2009; Petersson et al., 2012; as reviewed in Wilson et al., 2017). In relation to sequence learning, relatively specific roles have also been suggested for the TPJ, in the integration of both syntactic and lexical-semantic content in language (Price et al., 2016; Friederici, 2011, 2012), and in the integration of "what" and "when" encodings to facilitate sequencing rule generalisation (Orpella et al., 2020). Similarly specific roles have been suggested for the superior temporal gyrus, which has been implicated in both the ordered representation of categorical speech and/or phrase-level encodings, posteriorly (Yi et al., 2019; Chang et al., 2010), and in higher order lexical-semantic/syntactic integration, anteriorly (Brennan et al., 2012; Rogalsky & Hickok, 2009; Friederici et al., 2009).

From these findings, it appears that there are emerging points of agreement on the domain-general roles of specific brain regions in sequence processing. However, in order to clarify these roles, we must characterise dynamic neuronal responses at a suitably high temporal resolution.

Much of what we know about the time course of sequence processing thus derives from the results of electroencephalographic (EEG) and electrocorticographic (ECoG) studies, capable of measuring mega- and mesoscale neuronal population responses on the millisecond scales relevant to language (Beres, 2017). Intracranial or scalp EEG data can be averaged across trials to produce an *event-related potential* (ERP), describing time-locked responses to selected events (Sur & Sinha, 2009), interpreted as reflecting specific neurocognitive processes (Roach & Mathalon, 2008). A rich variety of EEG studies has revealed canonical components within ERPs elicited by manipulations of natural language and non-linguistic sequential stimuli. ERP components are stereotyped waveform features, characterised by commonalities in response latency, topographic distribution and polarity. Three components in particular have been associated with violations of sequence structure within and outwith language: the left anterior negativity, the N400, and the P600. These are discussed below.

The left anterior negativity (LAN) is a negative deflection in the ERP at around 300-500 ms elicited by certain classes of expectation violation (Gunter et al., 2000). Though topographically variable (Alemán Bañón & Rothman, 2019), the component is thought to emanate from left-hemispheric anterior regions. The LAN has been strongly associated with violations of morphosyntactic dependencies in natural language, including number and gender disagreement, verb tense violations, and subject-verb disagreement (Friederici & Meyer, 2004; Angrilli et al., 2002; Gunter et al., 1997; Münte et al., 1993; Kutas & Hillyard, 1983). In particular, it has been suggested that the LAN is most likely to be elicited by local (adjacent) dependency violations (Molinaro, Barber, et al., 2011; Molinaro, Vespignani, et al., 2011). A probable variant of the LAN, observed at a shorter latency of around 150-300 ms (known as the early left anterior negativity, or ELAN), has also been observed in cases where words in a sentence are from irreconcilable syntactic categories, such that it is impossible to build local phrase structure (Friederici, 2002). However, it has been suggested that the LAN does not specifically index the morphosyntactic processing of language, but rather represents broader proto-syntactic or morphemic expectation processing (Bornkessel-Schlesewsky & Schlesewsky, 2019). Consistent with this account, components with similar properties to the LAN have been observed in response to non-linguistic sequence structure violations, with sufficient exposure (Friederici et al., 2002; Hoen & Dominey, 2000).

More controversially, it has been suggested that the LAN does not in fact exist at all, but that it is in fact an epiphenomenon of two other components: the P600 and the N400 (Gonda et al.,

2020). The P600 (or *syntactic positive shift*, SPS) is characterised by a late positive deflection in the ERP (500-1000 ms latency) with a broadly centro-parietal topographic distribution, observed in response to a wide variety of expectation-driven syntactic anomalies in natural language (Gouvea et al., 2010; Friederici et al., 1993; Hagoort et al., 1993; Neville et al., 1991; Osterhout & Holcomb, 1992). The P600 has also been observed in response to dependency violations in AGL tasks (Silva et al., 2017; Tabullo et al., 2013; Christiansen et al., 2012; Lelekov-Boissard & Dominey, 2002). It has been suggested that the P600 indexes retrieval and structure-building processes, which appear to explain its elicitation not just by syntactic anomalies, but also in the resolution of well-formed long-distance dependencies (Gouvea et al., 2010). By contrast, the N400 is a negative deflection arising at latencies of 250-600 ms (Kutas & Hillyard, 1989, 1980) with a centro-parietal topographical distribution (Kutas & Federmeier, 2011), conventionally held to index lexical retrieval and semantic integration processes (Delogu et al., 2019). However, it has recently been proposed that the N400 might conceivably be elicited by both semantic *and* syntactic violations, but that it is spuriously characterised as a LAN in some cases due to the interaction of a broadly distributed N400 and right-posterior P600 (Tanner, 2019; Tanner & Van Hell, 2014). Despite this suggestion, analysis of at least one large dataset (N = 80 participants) has demonstrated that the LAN is observable at the level of individual participants and trials, indexing syntactic agreement independently of the N400 (Caffarra et al., 2019). In summary, therefore, whilst these components (LAN, N400, P600) remain incompletely understood, and the source of much debate, all are of potential relevance to a complete account of sequence processing.

As well as analysing continuous recordings in the time domain using ERPs, neural responses can also be characterised over time in the frequency domain, through spectral decomposition of recorded signals into content at different frequencies (Makeig, 1993; Makeig et al., 2004). Activity in specific frequency bands is considered to reflect differential contributions of distinct oscillatory processes (Herrmann et al., 2014). The event-related spectral perturbation (ERSP; see Makeig, 1993) is a now-ubiquitous time-frequency measure that comprises the mean power of the electrophysiological spectral decomposition across trials, computed relative to defined baseline activity. This is distinct from the information provided by event-related potentials (ERPs), which reflect the mean time-domain signal over trials. Using ERPs, activity across trials can "cancel out" if the individual trial responses are not phase-aligned, but with ERSPs this is not the case.

Consequently, whilst some time resolution is sacrificed to resolve frequency-specific detail, ERSPs can reveal more about the brain's electrophysiological response by describing not only phase-locked activity, but also the time-locked, phase incoherent activity not captured by ERPs.

Prior EEG studies have revealed evidence of characteristic time-frequency responses to sequencing violations. In particular, violations of natural language syntax eliciting a P600 have been associated with a beta (~13-30 Hz) and alpha (~10-13 Hz) band decrease exhibiting a broad topographical distribution (Schneider & Maguire, 2018; Lewis et al., 2016; Kielar et al., 2014; Davidson & Indefrey, 2007). At least one study has also reported theta (~4-10 Hz) increases in conjunction with alpha decreases (Regel et al., 2014). Triangulating these and other observations during sentence comprehension, it has been posited that, during the presentation of linguistic content, words or phrases are repeatedly syntactically and semantically unified over time to build structure unless semantic or syntactic anomalies disrupt the process, and that successful unification is associated with frequency-band-specific power increases in primarily frontal regions of the frontotemporal language network, including IFG (Bastiaansen & Hagoort, 2015; Hagoort, 2013). Accordingly, some studies have demonstrated that power in implicated bands gradually increases during the course of error-free sentence presentation (Segaert et al., 2018; Bastiaansen et al., 2010; Bastiaansen & Hagoort, 2015). Moreover, by comparing EEG responses to correct sentences, syntactically correct but meaningless sentences, and random word lists, disruption of semantic unification has been associated with a relative decrease in gamma power (~30-150 Hz) relative to content that can be successfully unified, whilst unsuccessful syntactic unification has been associated with a relative decrease in beta power, also in line with previously reported studies (Bastiaansen & Hagoort, 2015; Bastiaansen et al., 2010). However, this clear frequency-specific division of functionality has not been universally replicated, with compelling syntactic effects also observed in the gamma band (for example, Nelson et al., 2017), and a number of EEG/MEG (magnetoencephalography) studies reporting alpha/beta decreases in response to semantic violations (Kielar et al., 2018, 2015, 2014; Wang et al., 2012; Luo et al., 2010). It has been suggested that these findings can be reconciled with a more general predictive coding account (Friston & Kiebel, 2009; Rao & Ballard, 1999) in which beta synchronisation indexes the maintenance of phrase-level predictions destined for lower-order regions, and gamma activity reflects prediction errors destined for higher-order regions (Lewis et al., 2015; Lewis & Bastiaansen, 2015; Prystauka & Lewis, 2019).

As has been seen, many of the reported studies on syntactic processing have utilised natural language sentences. However, there remains a paucity of electrophysiological data collected under controlled statistical learning studies in humans. Moreover, neural responses to sequences not only contain information on the very short timescales captured by EEG (Beres, 2017), but also differ considerably even between immediately adjacent regions of the brain, for example within the inferior frontal gyrus, where adjacent subregions seem to be functionally heterogeneous (Wilson et al., 2017; Friederici, 2011). This motivates the use of imaging techniques with a high spatiotemporal resolution, such as electrocorticography (ECoG), which can produce data on a scale of millimetres by milliseconds (≤ 10 mm, ≥ 5ms; Asano et al., 2005).

In the previous chapter, behavioural evidence was presented for simultaneous adjacent and non-adjacent dependency learning by a neurotypical cohort of healthy adults, using a novel auditory AGL paradigm, the *AxB* task. In this study, we extended this work by characterising behavioural and electrophysiological responses in 12 intracranially monitored neurosurgical patients undertaking the *AxB* task. By undertaking both time- and time-frequency domain analyses on the recorded intracranial data, we aimed to clarify neural responses to non-linguistic sequential stimuli within the frontotemporal language network, adding to the limited body of existing ECoG evidence. In particular, we aimed to take advantage of a reported linear correspondence between ensemble ECoG and EEG responses (Kaur et al., 2013; Krusienski & Shih, 2010) in order to make comparisons with the extensive body of existing EEG evidence on the nature of syntactic ERP components.

We formed a number of predictions of expected behavioural and electrophysiological outcomes under this task. Behaviourally, we predicted that, as in healthy adult participants (see Chapter 2), adult neurosurgical patients would exhibit sensitivity to violations of the *AxB* grammar. Because analysis in the healthy cohort had previously revealed that self-reported fatigue was associated with lower scores on the explicit grammaticality judgement task, we also predicted lower scores in the patient cohort following implantation as a result of the influences of the hospital environment and surgical/medical interventions. Irrespective of explicit measures of performance, however, we predicted that reaction time effects would reveal implicit sensitivity to the grammar even in low performers, as in the healthy cohort. Neurophysiologically, we predicted that the AxB task would elicit neural responses in regions across the frontotemporal language network. In consideration of prior findings showing neural sensitivity to syntax even in the absence of awareness

(Batterink & Neville, 2013), as well as our previously reported reaction time effects (see Chapter 2), we predicted that both ERSPs and ERPs would reveal implicit and automatic sensitivity to the sequence structure. In particular, we expected sequences to elicit approximations of canonical ERP components including left anterior negativity and the P600, modulated by sequence grammaticality, and detectable in regions including ventrolateral prefrontal cortex. We also expected that any detected LAN would be enhanced in the "adjacent" sequencing violations relative to the "non-adjacent" violations, in line with previous suggestions (Molinaro, Barber, et al., 2011; Molinaro, Vespignani, et al., 2011). Finally, when comparing responses in the time-frequency domain, we predicted that violations of sequence structure would elicit suppressed alpha/beta power in critical regions, relative to consistent sequences, and that high gamma power would conversely be relatively enhanced for violation sequences versus consistent sequences.

Behavioural findings in this study revealed that the task was extremely challenging for the neurosurgical cohort, even with the inclusion of additional staged exposure phases designed to support learning. Chance performance was observed at the group level on the grammaticality judgement task, and no significant differences were observed between mean reaction times under different experimental conditions. However, a sliding window analysis of individual patient sessions did show spells of good performance. Furthermore, despite limited behavioural confirmation, differences in the neural responses of specific regions were observed between the experimental conditions at the group level, chiefly observable in the time domain. Moreover, these effects corresponded to canonical LAN effects reported in the electroencephalographic (EEG) literature, suggesting automatic and implicit reactivity to the artificial grammar, despite the difficulty in obtaining unequivocal behavioural data. Across all trials, task-related responses were also observed in the time-frequency domain within critical regions of the frontotemporal language network. However, the time-frequency effects of the task were difficult to discern using classical univariate analyses, as exemplified by the difficulty in identifying effects even in the highest-performing participant. As a consequence, we conclude that a multivariate analytical approach, rather than univariate electrophysiological analysis, is a possible route to the clarification of fronto-temporal sequence learning mechanisms.

3.3 Methods

3.3.1 Participant recruitment and ethics

All participants were adult neurosurgical patients (n = 12 intracranially recorded under the task, 7 male, 5 female, ages 19–55, median 33 years; see Table 3.1) diagnosed with medically refractory epilepsy and undergoing chronic intracranial pre-surgical monitoring at University of Iowa Hospitals and Clinics (UIHC) to identify seizure foci suitable for resection. Research protocols were approved by the University of Iowa Institutional Review Board (IRB ID No.: 200112047) and National Institutes of Health. Participation in the study did not impact on clinical monitoring or management. Informed written consent was obtained from each subject prior to their participation in the study, and participants were free to withdraw consent at any time without any impact on clinical evaluation. Sessions were suspended for at least three hours if a seizure occurred, and only resumed if the participant was willing and alert. These ethical guidelines are consistent with European Ethical Guidelines (Helsinki Declaration and H2020 EU guidelines).

Human intracranial electrode implantation is purely clinically motivated, a fact which inevitably constrains cohort size in human ECoG studies. However, before the end of the study, and to roughly determine the feasibility of collecting sufficient data, an *a priori* power analysis (Faul et al., 2007) was conducted using *AxB* task data previously collected in the Newcastle cohort (see Chapter 2). Based on group performance on the final run of the task in Newcastle (n = 50, $\mu = 0.77$, $\sigma = 0.22$, yielding effect size d = 1.23), it was revealed that, for a study on any similarly behaving cohort to reach one-tailed nonparametric significance of performance versus chance at the group level (to $\alpha = .05$), the minimum required sample size would be approximately 8 participants (see also Appendix 2: Supplementary figures, Figure 0.1). This was ultimately exceeded with the inclusion of 12 patients in the study.

Prior to participation, each patient completed audiometric and neuropsychological evaluation. No participants were found to have deficits in hearing considered likely to impact on performance of the *AxB* task. Four participants exhibited mild impairments of auditory attention, working memory or verbal fluency but these were not considered likely to prevent successful performance of this task.

Table 3.1: **Cohort of neurosurgical participants**.

No.	1	2	3	4	5	6	7	8	9	10	11	12
Identifier	**394**	**403**	**405**	**416**	**423**	**425**	**429**	**430**	**434**	**442**	**456**	**460**
Age, ♂/♀	23M	55F	19M	32M	49M	52F	32F	28M	40F	34F	31M	52M
Handedness	R	R	R	R	R	L	R	R	R	R	L	R
Wada result	L	L	L	L	L	-	-	-	L	L	L	L

3.3.2 Task design and stimuli

The task design largely followed the procedure described in Chapter 2, using slightly different computer peripherals (see below) and with the insertion of two additional phases before the rest of the task began: passive *pair exposure* and active *pair learning assessment*. These phases were intended to augment the incremental learning process. In the passive pair exposure and pair learning assessment phases, participants were first presented with 5 sequences each of the two essential "paired" elements, $\mathbf{A_1B_1}$ or $\mathbf{A_2B_2}$, words which remain associated with each other in all later grammatical stimuli of the task. This phase was therefore designed to promote attendance to the most task-relevant features of the stimuli. Furthermore, understanding this simple relationship would allow participants to most effectively extract the dependencies from subsequent sequences and thus potentially enhance later performance. During assessment, participants were tested on their ability to identify correct and incorrect **AB** pairings in 12 stereotyped trials. These contained the following patterns in the following order: $\mathbf{A_1B_1}$ (correct), $\mathbf{A_1B_2}$ (incorrect pairing), $\mathbf{A_1B_1}$ (correct), $\mathbf{B_1A_1}$ (incorrect order), $\mathbf{B_2A_1}$ (incorrect order and pairing), $\mathbf{A_1B_1}$ (correct), $\mathbf{A_2B_2}$ (correct), $\mathbf{A_2B_1}$ (incorrect pairing), $\mathbf{A_2B_2}$ (correct), $\mathbf{B_2A_2}$ (incorrect order), $\mathbf{B_1A_2}$ (incorrect order and pairing), $\mathbf{A_2B_2}$ (correct). This stereotyped ordering was designed to appear random whilst avoiding random clustering of conditions, to avoid impeding learning. If a participant failed to reach criterion performance (correct responses to at least 9 of the 12 trials), the pair exposure and assessment were repeated once more before proceeding to the usual exposure and test phases.

The task was run in an electrically shielded suite within the Clinical Research Unit of UIHC by members of the Human Brain Research Laboratory (HBRL) and, on three occasions, by the author. During performance of the task, which took 25–40 minutes, the door was closed, lights were

dimmed and the room was quiet except for stimulus playback. Playback was through two free-field speakers positioned 1 metre from the participant bilaterally. Forced-choice keyboard responses were provided through a handheld two-button response box. Participants carried out each phase of this task whilst reclined in a hospital bed, after receiving scripted verbal instruction from the researcher running the protocol (who remained within the suite, providing no task direction until the trials of each phase were complete). Following completion of each phase, the participant was queried on their willingness to continue with the task, and their wakefulness was assessed. Before and after the entire task, the participant's wakefulness was again assessed and they were queried on their willingness to continue with other experimental protocols.

3.3.3 Electrode configuration and acquisition

Participants were implanted with clinically indicated subdural ECoG arrays and depth electrodes. Electrode arrays were manufactured by Ad-Tech Medical (Racine, WI, USA). Depth electrode arrays each comprised 4-8 macro contacts with 5-10 mm spacing. ECoG arrays each comprised a grid or strip of platinum-iridium disc electrodes (2.3 mm exposed diameter, 5-10 mm spacing) encapsulated by a silicon membrane. Recordings were referenced to a subgaleal strip electrode. Of the cohort of n = 12 chronically implanted participants, 7 had left-hemispheric electrode coverage, 4 had right-hemispheric coverage, and one bilateral (see Figure 3.1). Across the cohort, 9 of the participants had previously demonstrated clear left-hemispheric language dominance during an intracarotid sodium amobarbital procedure (IAP; Wada, 1949) conducted by the clinical team, whilst of the remaining three participants without IAP results, two were presumed left-dominant on the basis of right-handedness, and one was of unknown status, being left-handed (Knecht et al., 2000; again see Table 3.1). Overall, there was coverage of the language-dominant hemisphere only in 6 of the 11 left-dominant participants, of the non-dominant hemisphere only in 4 participants, and of both hemispheres in one participant.

Local Field Potential (LFP) data was acquired with a Neuralynx Atlas system (Neuralynx, Bozeman, MT, USA), amplified, hardware bandpass filtered to between 0.1 Hz and 500 Hz (5 dB/octave roll-off), and digitised at a sample rate of 2000 Hz. A room microphone and presentation system auxiliary sound output were both also simultaneously recorded (at a minimum

of 16000 Hz, largely filtered 0–4000 Hz[*]) using the Neuralynx system, along with digital events generated by the MATLAB task script. Behavioural data including the pattern of responses for each participant was also stored by the task script for offline analysis.

Electrode locations were confirmed in each patient by the clinical and HBRL team via co-registration of pre- and post-implantation structural imaging supported by intraoperative photography (Nourski & Howard, 2015). Recording sites were linearly co-registered to the MNI152 T1 standard brain, placing them into MNI coordinate space. All recording sites were subsequently assigned to standard anatomical regions of interest (ROIs) via anatomical reconstruction of electrode location. This was aided in the case of subdural arrays by automated parcellation of cortical gyri (Destrieux et al., 2010, 2017) using the FreeSurfer image analysis software suite (http://surfer.nmr.mgh.harvard.edu/), and in the case of depth electrodes with reference to structural MRI sections.

Electrode recordings from sites implicated in seizure activity were excluded from subsequent electrophysiological analysis. Additionally, for the purposes of the study reported here and in Chapter 5, only sites of either hemisphere assigned to the ROIs listed in Table 3.2 below were analysed. These sites were primarily included on the basis of known relation to the auditory or frontotemporal-parietal language systems.

[*] In one case sound input was recorded at 32 kHz with filter cut-offs of 0.1-8 kHz; in another at 24 kHz and filter cut-offs of 0-6 kHz. Sound input channel data was used by the experimenter for validation of event alignment only and all settings were beyond adequate for this purpose.

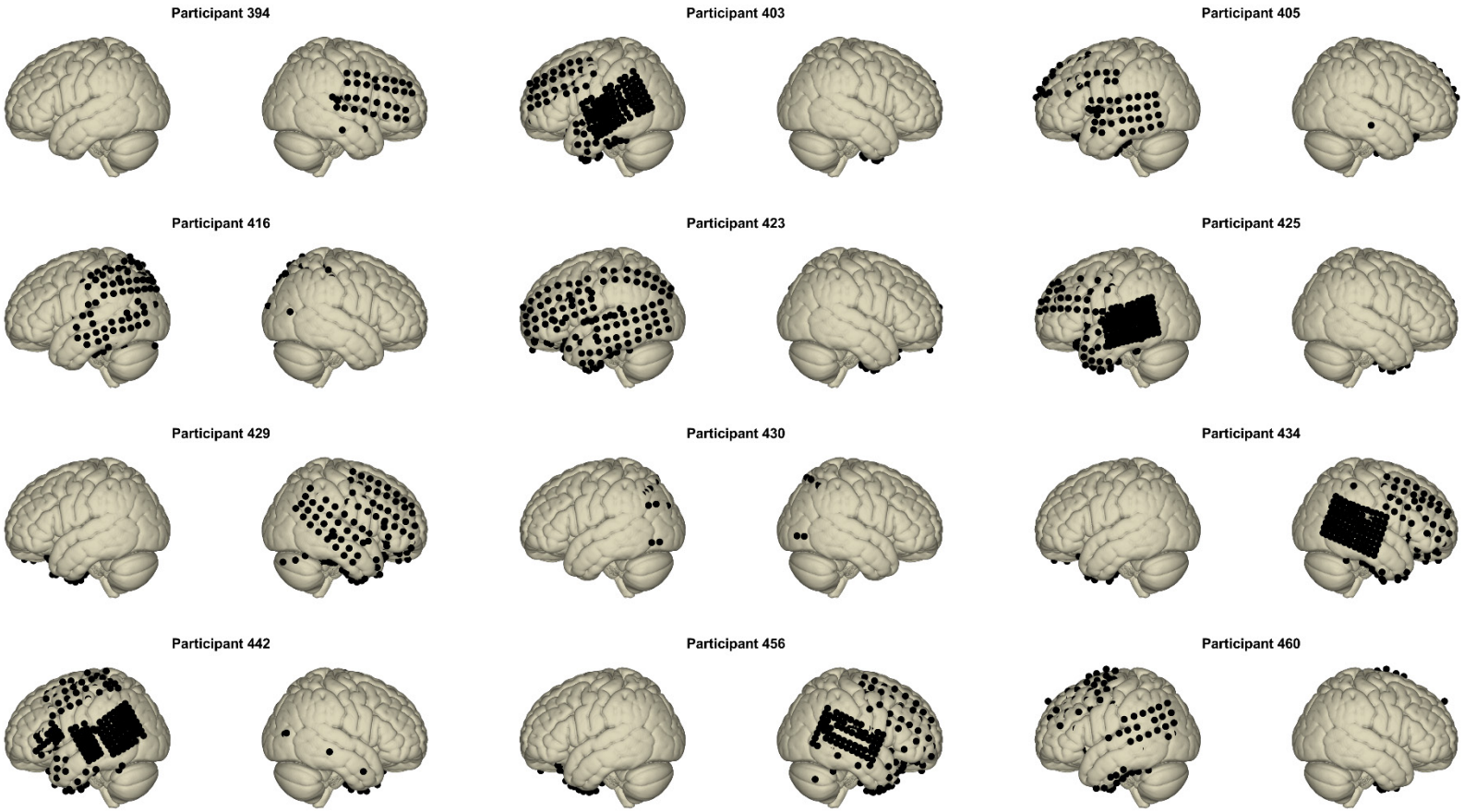


Figure 3.1: **Surface electrode coverage across the implanted cohort (n = 12).** Note that participant 430 had only stereoelectroencephalography (sEEG) electrodes implanted, rather than a subdural ECoG array as in the other participants.

Table 3.2: **Analysed regions of interest (ROIs), channel/subject coverage and centroids of channels in each region in MNI space across the cohort (n = 12)**

ROI name	№ channels (subjects)	ROI centroid left (x, y, z)	ROI centroid right (x, y, z)
Heschl's gyrus (HG)	L: 62 (6), R: 29 (4)	-44.9, -17.4, 5.9	46.6, -16.7, 3.9
Superior temporal gyrus (STG)	L: 147 (7), R: 69 (4)	-65.8, -19.9, 2.9	68.2, -16.0, 3.1
Superior temporal sulcus (STS)	L: 7 (4), R: 17 (4)	-54.8, -16.2, -12.8	56.1, -13.8, -9.2
Middle temporal gyrus (MTG)	L: 160 (8), R: 104 (8)	-65.7, -32.1, -6.7	66.0, -32.8, -7.0
Inferior temporal gyrus (ITG)	L: 70 (8), R: 24 (5)	-52.0, -29.7, -27.3	49.3, -22.3, -33.2
Inferior frontal gyrus (IFG) pars opercularis	L: 24 (6), R: 9 (4)	-58.9, 14.9, 15.3	62.1, 15.2, 12.5
Inferior frontal gyrus (IFG) pars triangularis	L: 33 (4), R: 18 (4)	-54.0, 26.8, 9.3	55.1, 30.8, 10.5
Inferior frontal gyrus (IFG) pars orbitalis	L: 11 (3), R: 11 (4)	-47.0, 36.5, -5.4	50.3, 31.9, -5.1
Frontal operculum	L: 3 (2), R: 1 (1)	-34.7, 19.6, -2.3	26.2, 16.5, -16.7
Insula	L: 18 (6), R: 14 (4)	-32.8, -3.4, 2.0	36.5, -5.0, -1.3
Hippocampus	L: 7 (4), R: 14 (6)	-32.1, -26.6, -11.5	25.1, -17.6, -16.1
Parahippocampal gyrus (PHG)	L: 22 (6), R: 18 (6)	-27.3, -18.8, -24.9	24.6, -17.9, -27.3
Angular gyrus (AG)	L: 37 (5), R: 27 (4)	-52.4, -66.3, 34.5	60.2, -55.1, 28.4
Supramarginal gyrus (SMG)	L: 70 (8), R: 41 (5)	-62.7, -42.4, 30.5	65.3, -36.6, 27.8
Middle frontal gyrus (MFG)	L: 110 (6), R: 67 (4)	-41.0, 29.4, 35.5	45.9, 35.2, 31.2
Orbital gyrus	L: 66 (8), R: 71 (9)	-26.1, 32.1, -15.4	29.9, 32.6, -17.1
Precentral gyrus	L: 46 (6), R: 38 (4)	-58.4, -3.8, 31.8	57.9, 0.0, 32.8
Postcentral gyrus	L: 38 (7), R: 21 (5)	-60.4, -17.1, 33.9	61.8, -12.7, 28.4

3.3.4 Pre-processing of channel recordings

Electrophysiological data were analysed offline using a custom MATLAB library with support from the EEGLAB (version 14.1.2; Delorme & Makeig, 2004) and FieldTrip (Oostenveld et al., 2011) toolboxes. Neuralynx data were first imported into MATLAB, and ECoG/LFP data and line noise removed using the demodulated band transform (DBT; see Kovach & Gander, 2016). Data was downsampled to 500 Hz using MATLAB's "resample" function (which applied to the data a finite impulse response, antialiasing lowpass filter with delay compensation) and further denoised by discarding the first principal component of the singular value decomposition (SVD) of the highpass-filtered signal (cutoff 160 Hz; see also methods of Kumar et al., 2020). To minimise jitter, recorded digital events were aligned with the auxiliary sound input (down to the per-sample level at 2 kHz) by shifting them using a custom event alignment script. Any data fragmented across multiple recording files was concatenated into one session, and continuous LFP data from selected channels initially epoched from -1500 to 4000 milliseconds around stimulus onset. For all analyses reported in this chapter, and most subsequent analyses (see Chapter 5), epochs were then truncated in order to omit the period following the earliest recorded button press across the cohort (which occurred at ~3100 ms post-stimulus-onset), and up to 100 ms prior. This was intended to exclude from analysis most activity relating to the overt motor response. This yielded epochs windowed from -1500 to 3000 milliseconds around stimulus onset.

3.3.5 Evoked potential analysis

Event-related potentials (ERPs), or evoked potentials, are derived from trial averaging of time-locked local field potential (LFP) data, and reflect coherent, replicable mega-scale neural responses to specific endogenous or exogenous events (Davis, 1939; Luck & Kappenman, 2013). Though typically associated with electroencephalographic (EEG) analyses, ERPs can be generated from ECoG signals in the same manner. However, the non-invasive nature of EEG has made it an accessible and prevalent technique in the cognitive neuroscientific literature and comparison with this data can yield useful insights; to this end, ERPs were computed from ECoG and depth electrode data at a regional level to support comparison with the EEG literature (Kaur et al., 2013; Krusienski & Shih, 2010). These large-scale ERPs were computed from the meso-scale intracranial data by averaging time-domain data across trials and within individual regions of interest. The purposes of the ERP analysis were thus twofold: 1) to reveal time-domain differences between responses to

"violation" and "consistent" sequences with a high temporal resolution; and 2) to determine points of correspondence between responses to this task and those reported in the existing EEG literature.

To calculate ERPs and compute contrasts between conditions, DBT- and SVD-denoised continuous channel data were bandpass filtered to retain signal between 0.5 and 100 Hz using a windowed sinc finite impulse response filter with very sharp roll-off (Kaiser window, HP: ~50 dB/octave roll-off; LP: >100 dB/octave roll-off). Continuous data were then split into epochs from -1500 to 3000 ms around stimulus sequence onset. Each epoch was baseline-corrected by subtracting its baseline mean. Baseline was taken to be 1350 to 1500 ms post-sequence-onset (corresponding to the inter-stimulus interval just prior to the third syllable, -150 to 0 ms relative to syllable onset). The resultant per-trial signals were then averaged across all trials, and under each condition of interest, to produce an "all-trial" ERP as well as ERPs for each of the conditions separately (for example, the "violation" and "consistent" conditions). ERPs from channels previously assigned to the same anatomical region of interest (ROI) according to the Destrieux parcellation algorithm (Destrieux et al., 2010, 2017) were then averaged together across the cohort to produce regional group mean ERPs. That is, data in a given region were summarised and statistically tested by aggregation of all electrodes across the cohort (as in Nourski et al., 2021; Haufe et al., 2018).This produced a weighted mean across participants. This effectively up-weighted those response profiles shared across electrodes in participants with the best regional coverage.

Before commencing statistical inference, and to provide a more appropriate timescale for comparison with the EEG literature, ERP data were windowed from 0 to 1000 ms around a point of reference 1500 ms into the trial (which was the time of the grammaticality manipulation, the onset time of the third syllable, and the end of the baseline period). Statistics were computed on this windowed signal and the baseline period only. Group mean ERPs under different conditions were subtracted from each other to produce a contrast. For example, "consistent" signals were subtracted from group mean "violation" signals to produce a difference wave comparing the amplitudes under both conditions. Cluster-corrected permutation testing (10 000 random permutations; or else the maximum number of unique permutations of the data, if smaller) was used to compute inferential statistics on this difference over time, assessing significance at $\alpha = .05$ (following the method of Maris & Oostenveld, 2007). This involved producing group-level surrogate distributions by permuting between conditions the trial-averaged data in electrodes from each subject (as motivated in Cohen,

2014; see also Theiler et al., 1992), always maintaining the dependence between electrodes of each participant. From these group-level surrogates, a null distribution of t-values was computed using a one sample t-test of the difference under the two surrogate conditions compared to 0, for each permutation. To determine two-tailed, cluster-corrected significance at each time point, the absolute t-values resulting from a dependent samples t-test on the true data (i.e. again, a one-sample t-test of the true difference under the two conditions against 0) were first thresholded according to the required α to produce supra-threshold clusters. For each observed cluster, the sum of the constituent t-values ("cluster mass") was computed. Likewise, the permutation distribution yielded, for each permutation, a set of clusters from which the absolute maximum cluster mass was recorded, producing a null distribution of maximum cluster masses. The *p*-value associated with each cluster in the observed data was finally derived from the proportion of permutations where the absolute maximum null cluster mass exceeded the absolute observed cluster mass (specifically: $p = (i + 1) / (N + 1)$, where i is the number of times the absolute maximum cluster mass exceeded the true cluster mass, and N is the number of permutations). These *p*-values, thus already cluster-corrected for multiple comparisons, were used to identify time points where neural responses differed significantly under each condition at the group level.

Note that the number of unique permutations always depends on the number of independent observations, obs, here equal to the number of participants with coverage in the given region. For a within-subjects study using dependent-samples testing, the number of unique permutations at the group level can be directly computed as 2^{obs} (see Maris & Oostenveld, 2007 for a comprehensive discussion). In this cohort, at the group level, some regions therefore yielded a limited number of unique permutations (see Table 3.2). Since $p = (i + 1) / (N + 1)$, the minimum achievable *p*-value is constrained in each region according to coverage. Accordingly, regional coverage was in some regions insufficient to allow α to be reached under any circumstances, and therefore insufficient to allow inferences to be made at the group level. These regions are marked in results plots with a dagger (†).

Prior to visualisation, ERPs were filtered once more to retain signal between 0.5 and 30 Hz using a windowed sinc finite impulse response filter with sharp roll-off (Kaiser window, HP: ~50 dB/octave roll-off; LP: >100 dB/octave roll-off).

Finally, as well as computing ROI-averaged ERP differences, per-contact ERP contrasts for every participant were also projected onto a three-dimensional standard brain model (Montreal Neurological Institute ICBM 152, Fonov et al., 2009, manually masked to retain only the cerebrum using BrainSuite 19b; Shattuck & Leahy, 2002). ERP difference waves were averaged over a specific time window for every contact in every participant, and these scalar values projected onto the cortical surface. This was performed using the NeuralAct toolbox for MATLAB (Kubanek & Schalk, 2015), which permits the creation of group-level cortical activity maps from ECoG by the following method. Firstly, locations of individual contacts (in MNI space) were projected onto the cortical surface with accuracy by inflating the brain model to produce a smooth convex hull, projecting from each contact along its normal to the hull surface, and then mapping this projection back to the original undulating cortical surface. Subsequently, activity at each vertex of the cortex was estimated by spatially convolving the value at each contact to approximate a localised spread of activity (here, using a 2D linear kernel fading to zero at 10 mm diameter), and then averaging values from all contributing contacts at every vertex. This yielded a three-dimensional brain showing a continuous spatial "heat map" of estimated ERSP differences across the cohort, despite the heterogeneous and discontinuous coverage inevitably associated with clinically placed electrodes.

3.3.6 Time-frequency analysis

Time-frequency analysis was undertaken to reveal frequency-specific responses at the group level. Mathematically, time-frequency analysis involves the decomposition of a single time-domain signal into multiple contributory time-domain signals of specific frequencies, resulting in an estimation of instantaneous power and phase of the original signal in a number of specific frequency bands over time (a process known as "spectral decomposition"). Spectral decomposition can be performed on any continuous data, including local field potentials recorded with ECoG or depth electrodes. In the resulting time-frequency map, activity in specific frequency bands is considered to reflect the contribution of distinct oscillatory processes in the brain (Herrmann et al., 2014).

The event-related spectral perturbation (ERSP; see Makeig, 1993) is a ubiquitous measure that comprises the mean power of the electrophysiological spectral decomposition across trials, computed relative to defined baseline activity. Time domain data from each channel was transformed into an ERSP map by way of a baselined Morlet wavelet transform (using the EEGLAB toolbox; Delorme & Makeig, 2004), to reveal frequency-specific power perturbations relative to the pre-stimulus

period. Decomposition was performed over 301 linearly spaced frequencies between 3 and 150 Hz, and over 100 equally spaced time points[2], to produce a 301-by-100 pixel map, normalised relative to a pre-sequence baseline beginning at -1500 ms and ending at 0 ms. Here, EEGLAB's "full" normalisation method was applied, which normalises first against per-trial baselines, and then again across trials to reduce the impact of individual noisy trials (Grandchamp & Delorme, 2011).

At the per-subject level, ERSPs were firstly produced across all trials ("all-trial ERSPs"), depicting the response relative to the pre-sequence baseline. ERSP maps were also produced for the subsets of "consistent" and "violation" trials. From these subsets, a per-subject "violation minus consistent" contrast was produced. This was a difference map depicting between-condition differences in response amplitude for every computed time and frequency. ERSPs under contrasted conditions were transformed prior to comparison so that they shared a common mean baseline. As for the ERP analysis, the calculation of group mean ERSPs (and statistical inference on between-condition contrasts) proceeded by aggregation of trial-averaged data across all electrodes in the cohort in a given region. Channels were pooled if falling within identical ROIs according to the Destrieux parcellation algorithm (Destrieux et al., 2010, 2017). As at the single-subject level, group-level ERSP maps were produced from all trials, as well as for subsets of "consistent" and "violation" trials. From these subsets, "violation minus consistent" contrasts were produced.

The procedures described above were also conducted for inter-trial phase coherence (ITC; see Tallon-Baudry et al., 1996) across participants. ITC quantifies the degree to which individual trial responses are phase-aligned at a given frequency and time. As with ERSPs, ITC values (range 0 to 1) were first calculated for each channel in each participant, then averaged over regional channels pooled across the cohort to form group-level all-trial ITC results.

[2] Using 2 cycles of input at 3 Hz, linearly increasing cycles at half the rate of frequency increase until 150 Hz. The default pad ratio of 2 increased the number of distinguishable frequencies to 301.

Also at the group level, exactly as with the ERPs, statistical significance of the ERSP differences was again determined by cluster-corrected permutation testing (10 000 random permutations; or else the maximum number of unique permutations, if smaller), assessing significance at $\alpha = .05$. That is, group-level surrogate distributions were computed by shuffling the ERSP maps from electrodes in each participant between conditions, always maintaining the dependence between electrodes. From these group-level surrogates, a null distribution of t-values was computed using a one sample t-test of the difference under the two surrogate conditions compared to 0, for each permutation. To determine two-tailed, cluster-corrected significance at each time-frequency point, the results of a dependent samples t-test on the true data (i.e. again, a one-sample t-test of the true difference under the two conditions against 0) were first thresholded according to the required α to produce 2-dimensional supra-threshold clusters in the time-frequency map. For each 2-D cluster, the sum of the constituent t-values ("cluster mass") was computed. Likewise, the permutation distribution yielded, for each permutation, a set of clusters from which the absolute maximum cluster mass was recorded, producing a null distribution of maximum cluster masses. The *p*-value associated with each cluster in the observed data was finally derived from the proportion of permutations where the absolute maximum null cluster mass exceeded the absolute cluster mass of the true cluster (specifically, $p = (i + 1) / (N + 1)$, where i is the number of times the absolute maximum cluster mass exceeded the true cluster mass, and N is the number of permutations).

Finally, statistical comparisons of high-gamma band power (taken to be 50-150 Hz) were also undertaken by an alternative method, averaging the spectral power recorded in the ERSP maps across the required frequency range to produce a set of time-resolved waveforms, and then statistically comparing them at the group level using cluster-corrected permutation testing ($\alpha = .05$), as previously described in 3.3.5 for the ERP waveforms.

3.4 Results

3.4.1 Behavioural analysis

Following implantation, participants undertook the task described in Chapter 2 (the *AxB* task, see 2.3.2). All n = 12 participants were actively tested, with one exception (participant 423, who was passively recorded whilst listening to the standard test sequences; see Table 3.3). Additionally, whilst the first 5 participants were only tested on the paradigm after implantation, low post-implantation task scores, reported below, meant that the subsequent 7 participants, as well as some unrecorded patients, were also assessed behaviourally on the *AxB* task before implantation. This allowed performance to be characterised in the absence of potential post-surgical confounds, and was intended to assist in prioritising those pre-surgical candidates exhibiting higher levels of performance as well as coverage prior to surgery. However, in practice, the mean pre-implantation score was generally low (overall $\mu = 0.57$, see Table 3.3), which gave the measure limited utility as a screening metric, and caused planned coverage of the language network to be the predominating factor.

Table 3.3 also shows a between-participant change to the AxB task. Whilst all subjects undertook multiple rounds of testing, each consisting of 48 trials, the two earliest participants experienced three test runs, balanced as described in 2.3.2 (144 trials rather than 96 trials across all runs). This was reduced to two runs in the remainder of participants to avoid patient fatigue.

Table 3.3: **Pre-implantation and post-implantation scores on the *AxB* task (n = 16).** All but the first 2 participants undertook 96 trials of testing in total across 2 runs. The latter 7 recorded participants (provided here with numeric IDs) were tested pre- as well as post-implantation. Participants assigned a letter rather than a number were assessed before implantation but not behaviourally assessed or intracranially recorded on the task after implantation, and were therefore excluded from further analysis. Exclusion mainly occurred in the event of insufficient coverage of relevant regions of interest. Additionally, at least one participant was excluded as a result of changes in their clinical condition. All but one recorded participants (*423*) were tested actively. Scores significantly different from chance indicated with an asterisk (* = $p < .05$, inverse binomial CDF, see 2.4.1).

Participant Number	Test runs	Pre-implant. First run score	Pre-implant. Last run score	Post-implant. First run score	Post-implant. Last run score
394	3	*Not tested*		0.42	0.40
403	3	*Not tested*		0.48	0.56
405	2	*Not tested*		0.42	0.46
416	2	*Not tested*		0.54	0.44
423	2	*Not tested*		*Passive only*	
425	2	0.46	0.58	0.46	0.48
429	2	**0.94***	**1.00***	0.44	0.50
430	2	0.54	0.44	**0.71***	**0.71***
434	2	0.46	0.54	0.56	0.46
(a)	2	0.44	0.52	*Not recorded*	
(b)	2	0.52	0.46	*Not recorded*	
442	2	0.58	0.50	**0.65***	0.50
456	2	0.52	0.52	0.56	0.44
(c)	2	0.52	0.56	*Not recorded*	
460	2	0.44	0.56	0.48	0.42
(d)	2	0.52	**1.00***	*Not recorded*	
		μ = 0.54, σ = 0.14	**μ = 0.61, σ = 0.20**	**μ = 0.52, σ = 0.095**	**μ = 0.49, σ = 0.086**

Pre- and post-implantation results are summarised in Table 3.3 and Figure 3.2. At the group level, pre- and post-implantation, performance on first and last runs of the task was not significantly different from chance (Wilcoxon signed rank, $\alpha = .05$). Pre- and post-implantation, there was no significant difference between first and last run performance (Pre-implantation: Wilcoxon signed rank, $\mu_1 = 0.54$, $\mu_2 = 0.61$, $p = .20$; Post-implantation: Wilcoxon signed rank, $\mu_1 = 0.52$, $\mu_2 = 0.49$, $p = .25$). Although mean performance on the final run was lower after implantation than before on those participants tested in both instances, this difference was not significant (Wilcoxon signed rank, $\mu_1 = 0.61$, $\mu_2 = 0.49$, $p = .31$). Unlike performance in the behavioural study reported in Chapter 2, neither pre-implantation nor post-implantation performance could be said to be truly bimodally distributed, but there were high performing outliers, notably two pre-implantation ceiling performers (429 and *(d)*, see Table 3.3 and Figure 3.2) and two patients who performed above chance in one or more of their post-implantation test runs (430 and 442).

Overall chance performance may not necessarily reflect a total inability to perform better than chance on the task across its entire duration. A chance score might instead reflect fluctuating high and low performance that simply averages to chance over time. With this in mind, a sliding window average of behavioural performance was computed on final run responses for each participant (see Figure 3.3). This suggested that performance fluctuated over time, and highlighted individual trials where sliding window performance breached chance levels, but no statistical inferences as to the likelihood of such breaches should be made from the analysis in its present form.

As in Chapter 2, performance was also broken down by experimental condition (Figure 3.4). Pre-implantation results showed significantly higher grammaticality judgement performance for consistent versus violation sequences, echoing a similar bias observed in Chapter 2 across the chance performers and overall cohort. Post-implantation results revealed no significant conditional bias (Wilcoxon signed-rank, $\alpha = .05$).

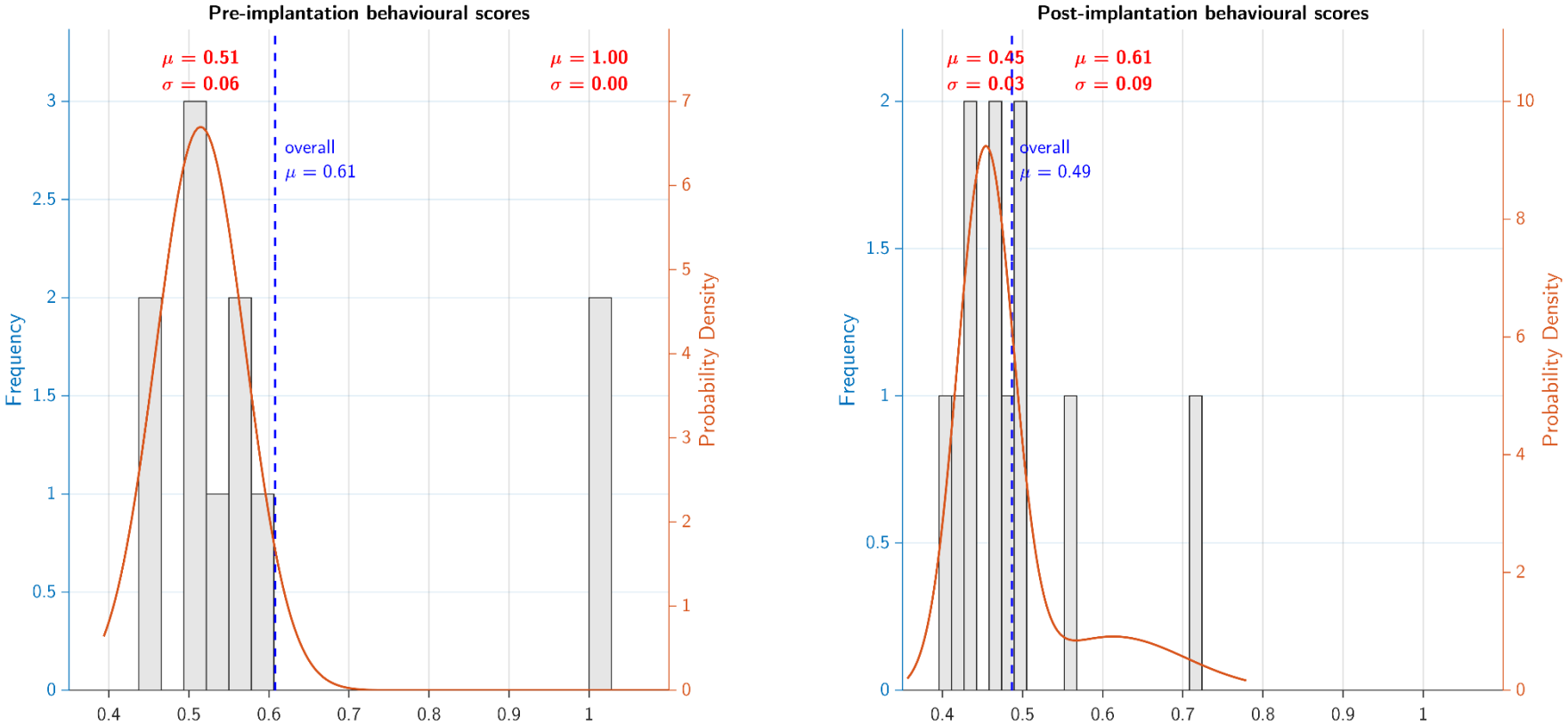


Figure 3.2: **Distributions of scores by pre- (n = 11) and post-implantation (n = 11) neurosurgical patients on the final run of the *AxB* task.** Note that these two groups, though the same size, do not contain identical participants. Some implanted participants were not recorded pre-implantation, and some participants assessed pre-implantation were not subsequently recorded during the paradigm (see main text and Table 3.3).

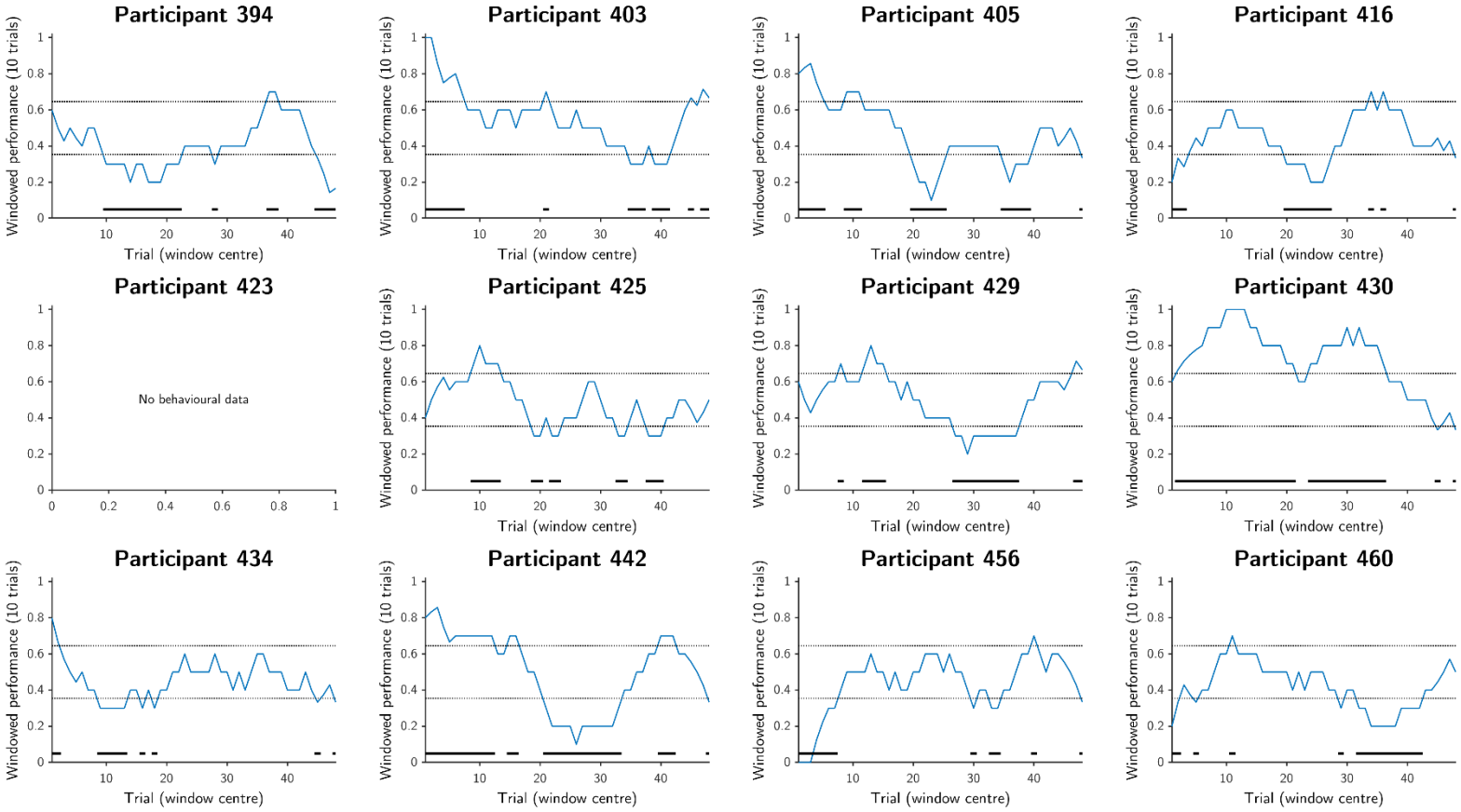


Figure 3.3: **Sliding window post-implantation performance of n = 12 neurosurgical patients on the *AxB* task, final run.** The trace in each plot shows sliding window final run performance over 10 trials. Limits (dotted lines) were computed from the inverse binomial cumulative distribution function and used to determine periods of potentially significant performance (black bars). Participant 423 was passively exposed to stimuli only. Windowed performance can be seen to fluctuate over time, often beyond the binomial limits, but note that no futher inferences should be made as to the significance of these periods.

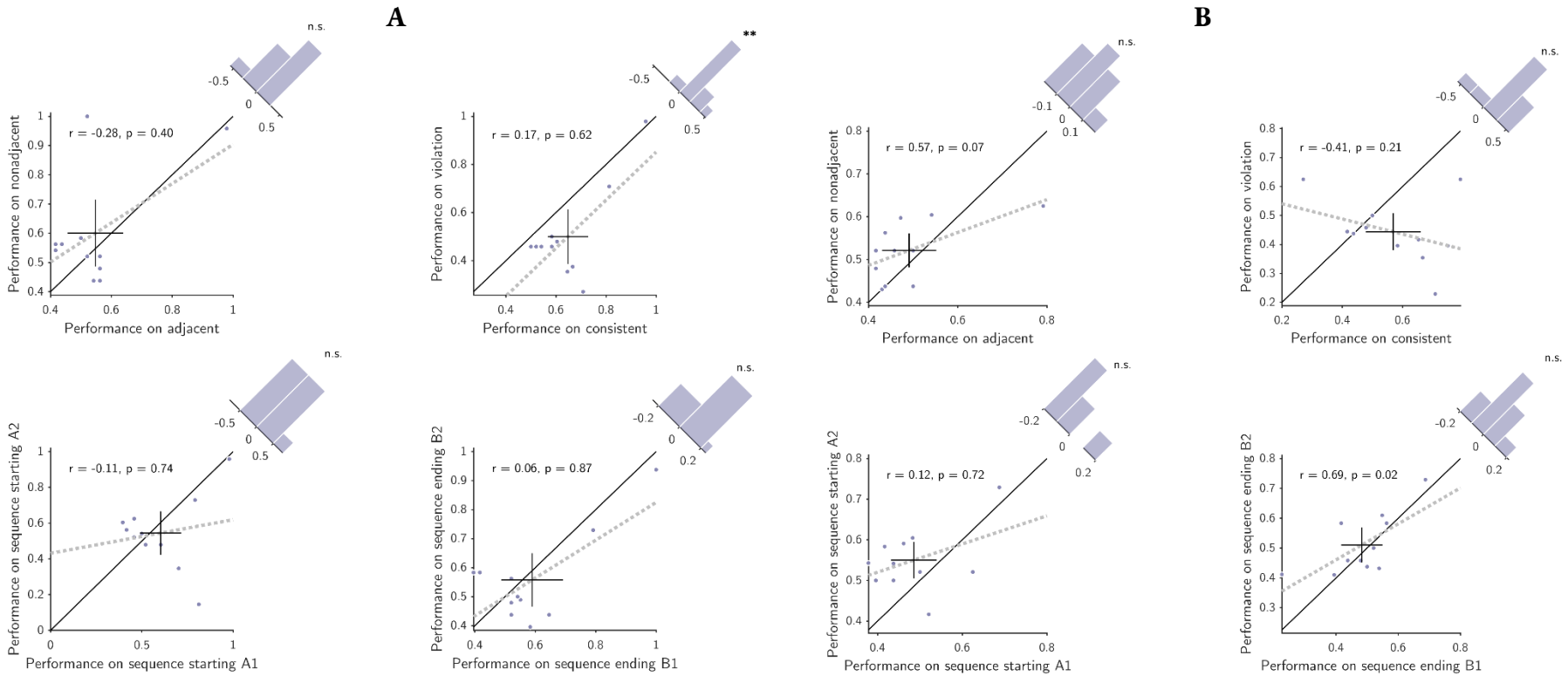


Figure 3.4: **Correlation and comparison of pre- and post-implantation performance by condition**. Panel A shows pre-implantation final run performance (n = 11) broken down by stimulus condition. Wilcoxon signed-rank revealed significantly higher performance on consistent versus violation sequences ($\mu_{cons} = 0.65$, $\mu_{vio} = 0.5$, $p = .0020$), as observed in results on the cohort reported in Chapter 2, but no other significant differences between performance on opposing conditions. Panel B shows the same analysis on post-implantation participants (n = 11). There were no significant differences between performance on opposing conditions. Spearman rank correlations revealed a significant positive relationship between scores on sequences ending in $\mathbf{B_1}$ and $\mathbf{B_2}$ ($r = 0.69$, $p = .017$). (* = $p < .05$, ** = $p < .01$)

Finally, since Chapter 2 demonstrated that grammaticality judgement in the *AxB* task appeared to be a measure of explicit awareness of the underlying grammar, an analysis was performed on reaction times collected under the post-implantation task, in order to reveal possible evidence of implicit sensitivity to the sequence structure (as also reported in Chapter 2). However, this analysis revealed no significant differences between mean per-participant reaction times under different conditions of the task (paired t-test on log-transformed reaction times, $\alpha = .05$; see Figure 3.5).

In conclusion, summary metrics of reaction times and performance on the grammaticality judgement task did not reveal significant sensitivity to grammaticality conditions in implanted participants. However, windowed performance was potentially suggestive of fluctuating responsiveness. In addition to this, the absence of significant overall performance or reaction time effects, especially across this multifactorially affected cohort, did not preclude the possibility of observing implicit or automatic sensitivity to the grammar in the neural data. Neural effects of grammaticality have previously been reported in the absence of awareness (Batterink & Neville, 2013). Thus, to provide evidence for syntactic sensitivity under the task, it was considered that an appropriate electrophysiological starting point would be an analysis well-suited to comparison with the canonical literature. As described in section 3.2, a number of electrophysiological responses to grammaticality have previously been characterised using electroencephalography, which motivated an evoked potential analysis on the ECoG dataset. This is described in the following section.

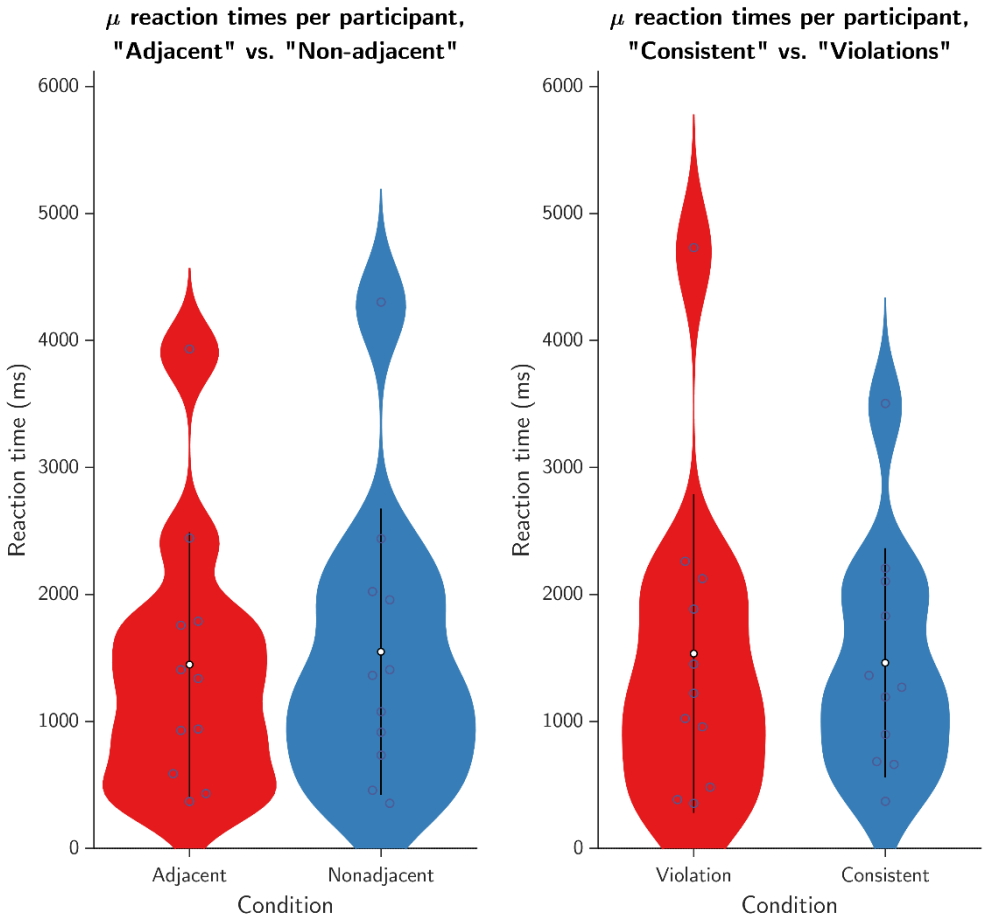


Figure 3.5: **Post-implantation reaction times on the deterministic task by structural condition (n = 11).** Shown are violin plots of reaction times (RTs) under the two manipulations of sequence structure. RTs were averaged across all trials of both testing runs in each participant. Each violin plot shows the individual data points, the probability density for each distribution, and is overlaid with the group mean and standard deviation. Group-level comparison of RTs revealed no significant differences between reaction times under each of the structural conditions (paired t-test on log-transformed data, $\alpha = .05$).

3.4.2 Evoked potential analysis

Figure 3.6 shows overall mean event-related potentials (ERP) across all trials. These were produced by averaging across every channel in a given region of interest (ROI) across the implanted cohort of n = 12 participants (see 3.3.5). Note that participant 423, whilst excluded from behavioural analyses as a passive participant, is incorporated into the following electrophysiological analyses. ERPs are here reported across a window of 1350 to 2500 ms, a period surrounding the final word of the sound sequence (the **B** element, which began at 1500 ms and ended at 2100 ms), as described in 3.3.5. The 1350 to 1500 ms pre-syllable baseline is also shown for reference. In this figure, the signal is shown *z*-scored relative to baseline (note that the vertical axis is flipped, as in canonical ERP plots). The *z*-scored signal was defined as $\boldsymbol{z} = \frac{\boldsymbol{x}-\bar{x}_{baseline}}{s_{baseline}}$, where x is the vector of raw samples, $\bar{x}_{baseline}$ and $s_{baseline}$ are the mean and standard deviation across the baseline period respectively, and $\boldsymbol{z}$ is the vector of *z*-transformed samples (see for example Tamalunas, 2021).

Responses deviating significantly from the pre-syllable baseline at the group level (surviving cluster correction for multiple comparisons) can be seen in selected left hemispheric regions including Heschl's gyrus, precentral gyrus and orbital gyrus. However, in a number of regions where coverage is insufficient to ever reach α at the group level using the permutation test (grey panels, marked with a dagger, "†"), for example right Heschl's gyrus, it can nevertheless be seen that the *z*-scores describe large shifts from baseline.

Figure 3.7 shows the group-level ERP contrast for "violation" minus "consistent" conditions across the cohort. ERPs are again reported across a window of 1350 to 2500 ms, around the time of the final monosyllabic sequence item, again including the pre-syllable baseline. The onset of the final word at 1500 ms marked the onset of the grammaticality manipulation. Note also that the design of the task ensured that the stimulus presented during this manipulation was acoustically matched between conditions (see 2.3.3). Among the regions assessed, no significant post-manipulation effects (surviving cluster correction for multiple comparisons) were observed bilaterally at the group level. Note, however, that more than half of the tested regions had insufficient coverage to support group level inference using the permutation test (see 3.3.5). Left IFG pars triangularis, opercularis and orbitalis all exhibited relatively early non-significant negative deflections at around 400-600 ms and

non-significant late positive deflections at around 800 ms post-onset, not unlike the canonical LAN and P600 responses, though no group-level statistical inferences can be drawn from these findings.

As previously described in 3.2, multiple studies have provided evidence that syntactically relevant ERP components may be modulated by factors including dependency distance, and a large body of evidence has also suggested that engagement of IFG is similarly complexity-dependent. Thus, although we could provide no evidence for overall grammaticality effects in IFG or other regions, we considered it appropriate to break down the grammaticality contrast into two distinct comparisons: a "violation minus consistent" contrast for adjacent (**xAB**) sequencing dependencies, and the equivalent comparison for non-adjacent (**AxB**) dependencies. Since ERP studies of natural language syntax typically compare responses to relatively specific syntactic manipulations, the aim in this case was to reveal effects obscured by the amalgamation of sequencing conditions in the overall contrast.

Figure 3.8 shows the group-level evoked potential grammaticality contrast specifically for left-hemispheric regions broken down by "adjacent" and "non-adjacent" sequencing conditions (i.e. grammaticality contrasts for **xAB** and **AxB** sequences) across the implanted cohort of n = 12 participants, using the same window of analysis as above (1350 to 2500 ms, including pre-syllable baseline). Results for right-hemispheric regions (which revealed no significant effects) are reported in Figure 0.2 and Figure 0.3, Appendix 2: Supplementary figures. In Figure 3.8, it can be seen that in all but one of the regions, neither of the "adjacent" and "non-adjacent" grammaticality contrasts revealed any significant effects at the group level using cluster-corrected permutation testing. However, the clear exception to this was left IFG pars opercularis, which exhibited a significant period of early negativity at the group level after cluster correction, under the "adjacent" grammaticality manipulation (from around 200-500 ms post-manipulation).

Under the "adjacent" contrast, IFG subregions all exhibited prominent non-significant relative negativity to violations between ~0 and 600 ms post-syllable-onset, whereas under the non-adjacent contrast, however, the same regions appeared to exhibit non-significant positivity to violations from ~0-1000 ms post-onset. Additionally, despite the differing direction of adjacent/non-adjacent effects, all IFG subregions exhibited negative-going deflections in both contrasts at ~400 ms, which appeared to produce the prominent, but non-significant, negativity observed in the overall grammaticality contrasts. Meanwhile, STG exhibited a slightly negative-going non-significant

deflection in all conditions, enhanced in the presence of a violation, resulting in a very early relative negativity (~0-100 ms) in both adjacent and non-adjacent contrasts. This subtle early negativity appears consistent with accounts of the auditory N100 response, which is a very early negativity that appears to emanate from superior temporal gyrus (Boutros et al., 2011) and is enhanced in response to unpredicted stimuli (Nordby et al., 1996; Schafer & Marcus, 1973).

Finally, in Figure 3.9, the "adjacent" and "non-adjacent" grammaticality ERP contrasts are shown projected onto a standard cortical surface, visible as relative positivities or negativities in field potential for the "violation" trials relative to "consistent" trials. Here, the unfiltered difference waves under the two contrasts have been averaged over two time windows: firstly from 0 to 250 ms post-syllable-onset (1500 to 1750 ms post-sequence-onset), and secondly from 250 to 500 ms post-syllable-onset (1750 to 2000 ms post-sequence-onset). In the "adjacent" grammaticality contrast, but not the "non-adjacent" equivalent, a notable (though weak) left-hemispheric negativity was localised to ventrolateral prefrontal and orbitofrontal cortex. This was most prominent at 250-500 ms post-onset but continued for at least another 250 ms. Strong relative negativity was visible in middle temporal gyrus and broader areas of the left temporal lobe in both periods of analysis, and in both "adjacent" and "non-adjacent" contrasts. Left hemisphere positivity in the "adjacent" contrast largely centered on superior temporal gyrus (short-lived, absent by 500-750 ms) and the anterior temporal lobe (long-lived, present at least between 250 and 750 ms, and not resolvable within the Destrieux-parcellated ROI analysis). In the "non-adjacent" contrast, however, weak positivity rather than negativity was observed in IFG at 250-500 ms, and a much stronger positivity was observed around the temporoparietal junction (TPJ) at 500-750 ms (which was not easily discernable from the ROI results).

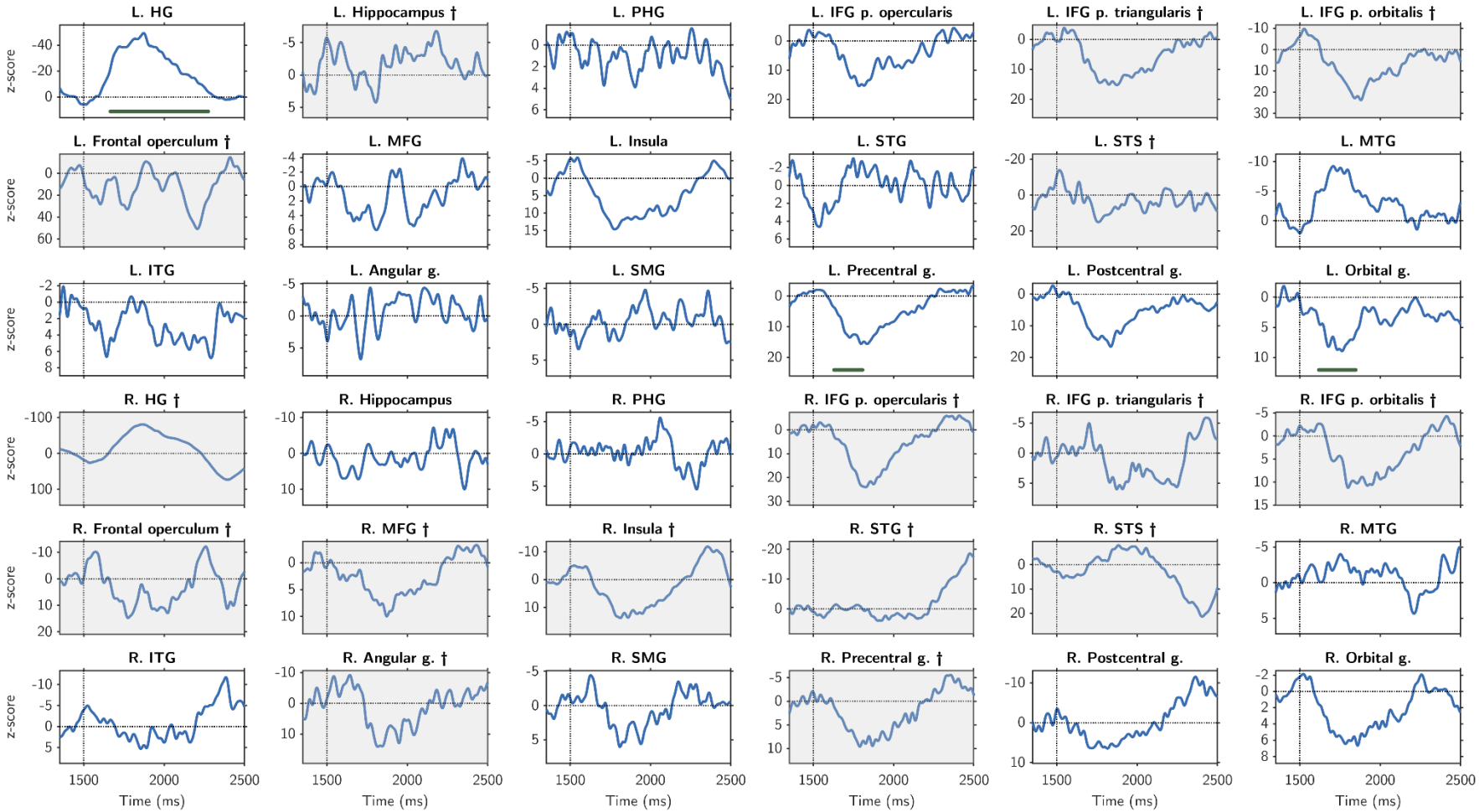


Figure 3.6: **Group mean (n = 12) event-related potentials across all trials.** Overall mean ERPs (pooling all channels) are shown for each region (1500-2500 ms), *z*-scored relative to baseline (1350 to 1500 ms). The transition between the end of the baseline and the onset of the third syllable is marked by a vertical line. Solid lines depict ERPs. Significant time points after cluster-corrected permutation testing ($\alpha = .05$; $\ll$10 000 iterations) are indicated with green dots. Significant responses can be seen in selected regions of the left hemisphere. Grey plots marked with a dagger (†) depict regions with insufficient group-level coverage to reach α in permutation testing. *Z*-scores still show large shifts from baseline in many such regions.

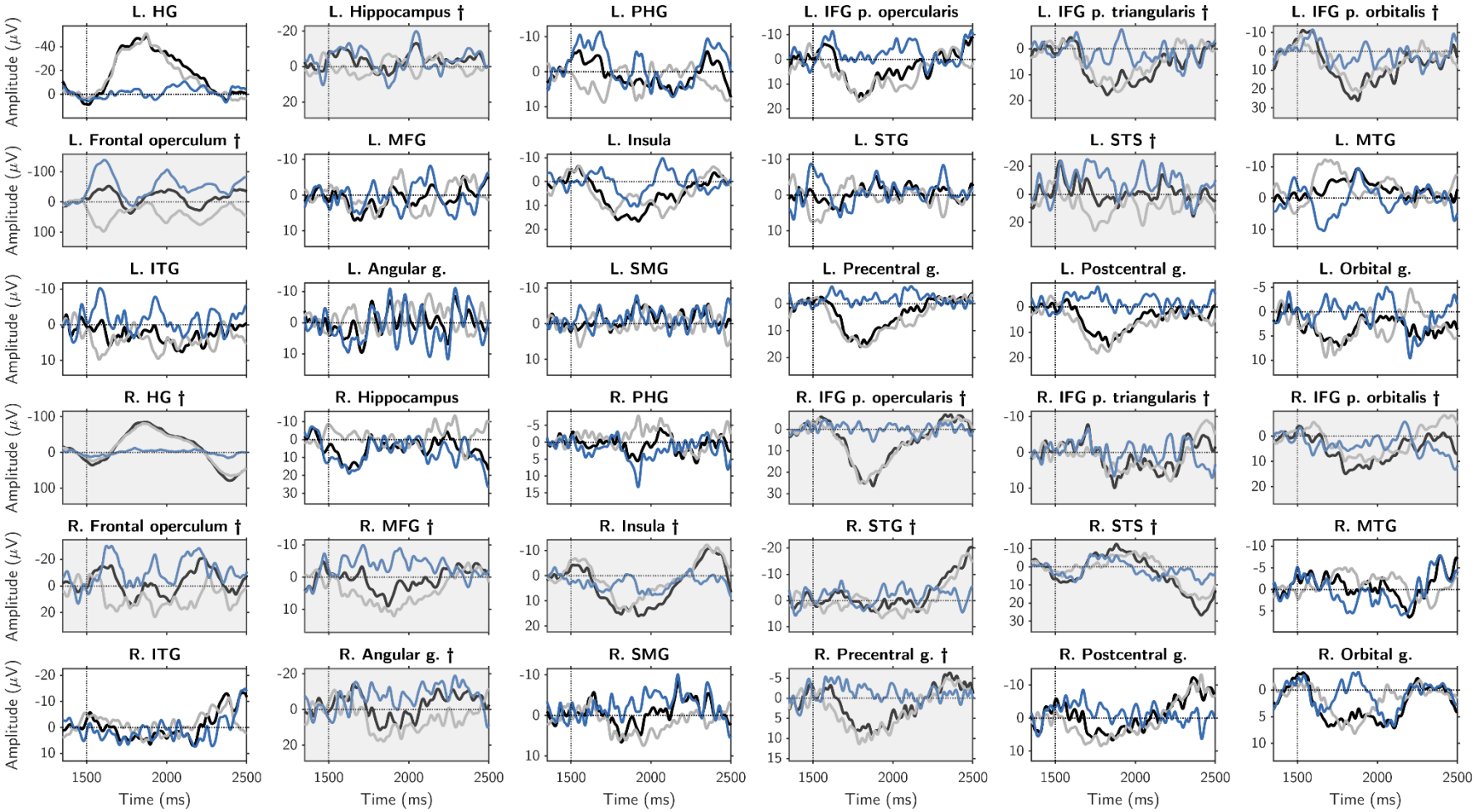


Figure 3.7: **Overall group mean (n = 12) evoked potential contrasts under "violation" minus "consistent" conditions of the AG.** Overall mean (equivalent to pooling channels across participants) and difference wave shown for each region and condition. Black trace: violation; Grey trace: consistent; Blue trace: violation minus consistent. No significant time points were revealed by cluster-corrected permutation testing ($\alpha = .05$; $\ll$10 000 iterations). Grey plots marked with a dagger (†) depict regions with insufficient group-level coverage to reach α in permutation testing.

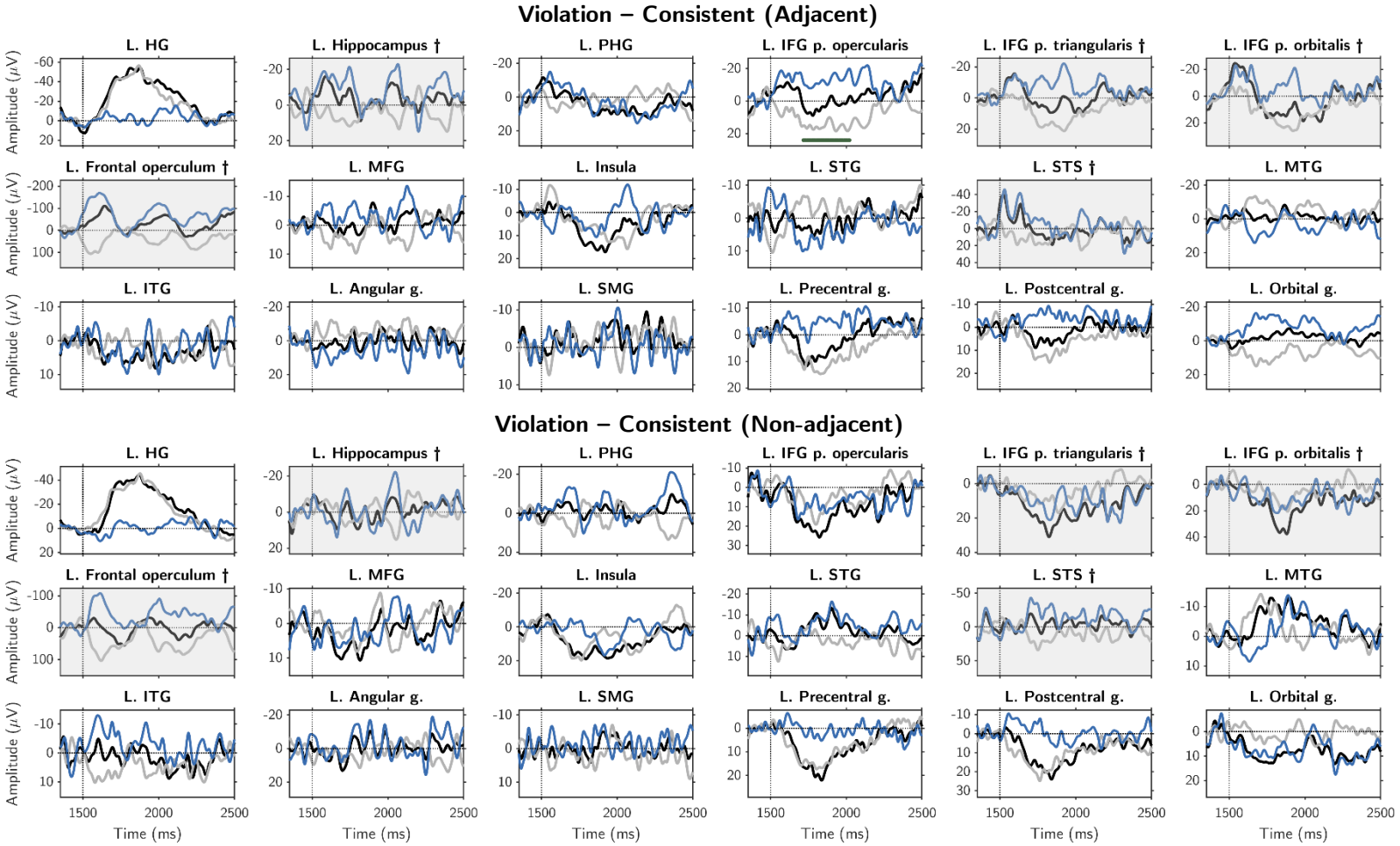


Figure 3.8: **Group mean (n = 12) evoked potential "violation" minus "consistent" contrasts for "non-adjacent" and "adjacent" sequences separately (left hemisphere only).** Overall mean and difference wave shown for each condition. Black trace: violation; Grey trace: consistent; Blue trace: Violation minus consistent. Significant time points revealed by cluster-corrected permutation testing (α = .05, $\ll$10 000 iterations) indicated with green dots. Grey plots marked with a dagger (†) depict regions with insufficient group-level coverage to reach α in permutation testing. Under the adjacent grammaticality contrast, left IFG pars opercularis showed significant early negativity at the group level.

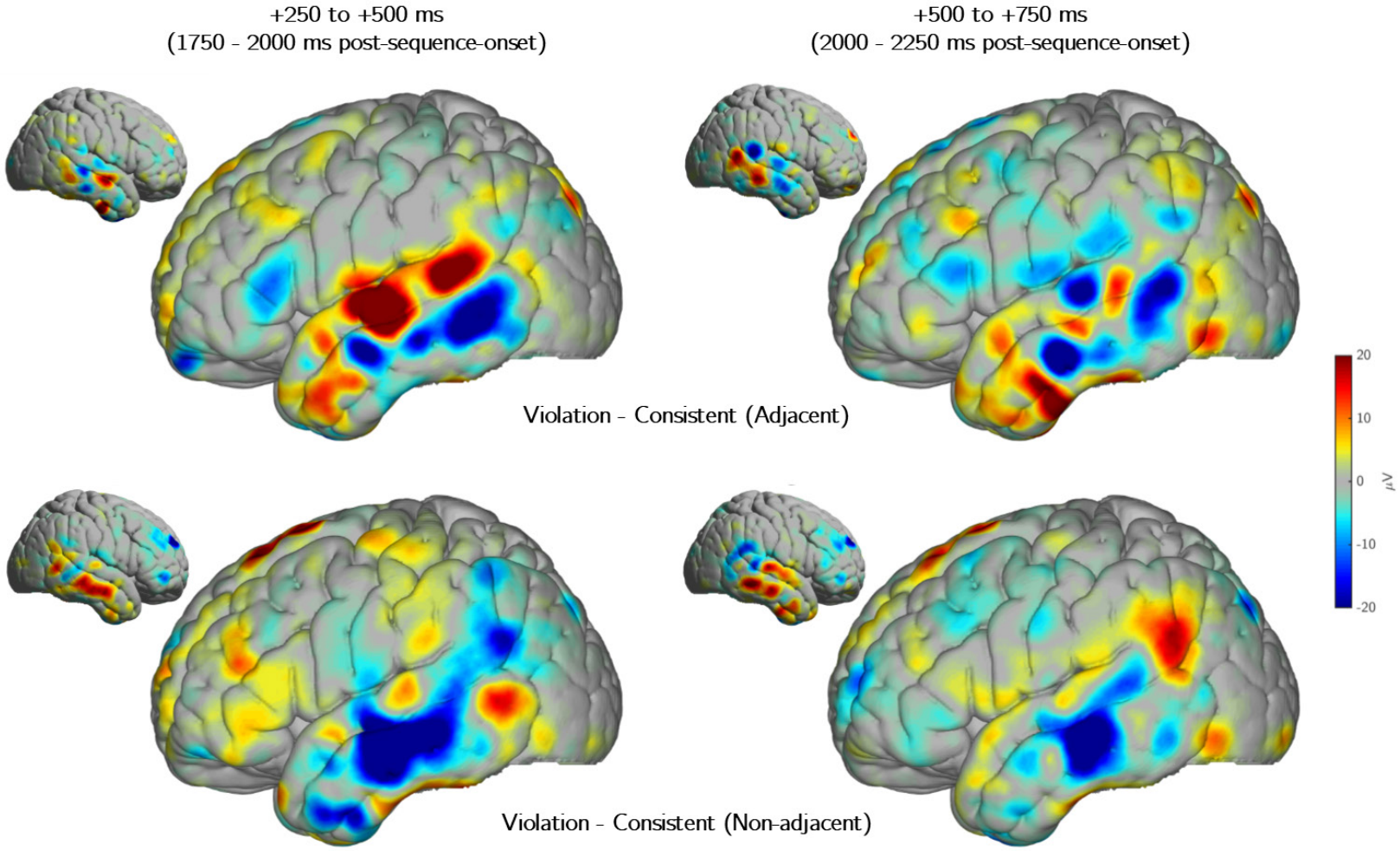


Figure 3.9: **Cortical projections of group-level (n = 12) grammaticality ERP contrasts.** Each brain shows ERP differences projected to a standard brain model on an electrode-by-electrode basis, across the pooled electrodes of the entire cohort. ERP differences were averaged within specific time windows prior to projection (250-250 ms post-syllable-onset; 500-750 ms post-syllable onset).

In summary, the group-level ECoG/depth electrode ERP analysis suggested significant sensitivity to a specific class of syntactic manipulations in IFG pars opercularis, a key region of the frontotemporal-parietal language network (discussed further in section 3.5.2). A number of ERPs featured deflections seemingly consistent with canonical EEG components including the N100 and syntactic P600/LAN complex, though such deflections were non-significant in all but the aforementioned region of interest (ROI), IFG pars opercularis, where significant early negativity was observed in the grammaticality contrast for adjacent sequences. From non-significance-masked cortical projections, late positivity was apparent in many regions, especially the anterior temporal lobe in the adjacent case and the left temporoparietal junction (TPJ) in the non-adjacent case. Meanwhile, from the same projections, it appeared that early negativity was generally prominent in the middle temporal gyrus and TPJ, but additionally appeared to emanate weakly from left inferior frontal gyrus in the adjacent case, in general correspondence with the one significant group-level ROI contrast. We also saw signs of widely distributed negativity within the temporal lobe, but there was limited evidence either here or elsewhere of centro-parietal negativities that might correspond to the timing of an N400 response, in line with prevalent accounts describing a syntactic LAN/P600 response and a separate, primarily semantic N400 response that would not have been elicited by our nonsense stimuli.

3.4.3 Time-frequency analysis

As described in "Methods" (see 3.3.6), spectral decomposition was undertaken for each epoch, in each channel, in every region of interest (ROI). Spectral content was averaged across selected epochs and normalised with respect to a baseline to produce an event-related spectral perturbation (ERSP) map for every channel. The group-level mean for each ROI was computed across all channels assigned to a given region, pooled across the cohort. This was performed across all trials as well as under individual conditions.

Figure 3.10 shows the all-trial group-level ERSP results for every ROI, normalised against the pre-sequence baseline. This effectively produces a "post- versus pre-stimulus" contrast in every region. Red-shifted colours in the map are indicative of relative enhancement of power relative to the baseline, whereas blue-shifted colours denote relative suppression. It can immediately be seen that the sequences elicited stimulus driven responses in a number of regions, manifesting as broadband responses to each monosyllable of the three item sequences (visible as vertical stripes in the spectral

maps). The strongest power perturbations were observed in bilateral Heschl's gyri, superior temporal gyri and right superior temporal sulcus, dominated by relative gamma-band (~30-150 Hz) enhancement and relative low-frequency suppression. However, stimulus-driven responses were also observed in a number of other regions across the left-lateralised frontotemporal network, and in corresponding regions of the right hemisphere. In the right hemisphere, these primarily manifested as gamma-band enhancement aligned with the presentation of each syllable, for example in right IFG pars opercularis and triangularis, insula, angular gyrus, supramarginal gyrus, and precentral and postcentral gyri. In the left hemisphere, the stimulus-driven response was dominated by relative low-frequency suppression spanning the delta (~0.5-4 Hz), theta (~4-10 Hz), alpha (10-13 Hz) and beta (13-30 Hz) bands during and/or after sequence presentation. This effect was observed prominently during sequence presentation in left IFG pars opercularis and triangularis, and precentral and postcentral gyri. However, a similar effect was observed around or following cessation of the sequence at 2100 ms in left IFG pars orbitalis, middle frontal gyrus and supramarginal gyrus.

In summary, the all-trial ERSP results are consistent with the all-trial ERP results, demonstrating clear shifts from baseline in a number of critical regions of the auditory and frontotemporal language network. However, they additionally depict distinct patterns of relative high-frequency enhancement and low-frequency suppression during and following sequence presentation.

Figure 3.11 shows the group-level all-trial inter-trial coherence (ITC) results for frequencies below 40 Hz (the approximate maximum frequency at which non-zero coherence could be identified at the group level). As with the ERSP results, the group-level mean ITC for each ROI was computed across all channels assigned to a given region, pooled across the cohort. At low frequencies, in a number of critical auditory- and language-relevant regions, increased coherence can be seen correspondent with the presentation of one or more individual syllables of the auditory sequences. This includes prominent effects in bilateral Heschl's gyri and bilateral IFG pars opercularis, bilateral IFG pars orbitalis, bilateral precentral and postcentral gyri, left IFG pars triangularis, right angular gyrus and right supramarginal gyrus. Some regions, including bilateral Heschl's gyri and bilateral IFG pars opercularis, also appeared to show a fourth brief post-sequence increase in coherence that resembled that of the preceding three syllabic responses, potentially representing an auditory steady-state response to the rhythmic syllabic presentation (Manting et al., 2021).

Finally, contrasts were also computed on the ERSP maps and high gamma power (see 3.3.6). Figure 3.12 shows these contrasts in specific regions, both across the high gamma band only (50–150 Hz, middle column) and the ERSP as a whole (right column), alongside auditory cortical high gamma power in the same hemisphere for reference (left column). As with the ERP contrasts, these identified very few regions exhibiting significant differences in responses between conditions at the group level. As a consequence, only a subset of the results are shown here, including those regions showing effects (see Figure 3.12), whilst the rest are included in Appendix 2: Supplementary figures (Figure 0.4 to Figure 0.9).

The one significant result under the overall grammaticality (violation minus consistent) contrast is shown on the top row of Figure 3.12. Under the ERSP comparison (rightmost column), this revealed subtle relative suppression in the right middle temporal gyrus (MTG) between 2000 and 3000 ms at around 30-50 Hz. No effects were identified across the high gamma band at the group level (middle column). As in earlier analyses, we also broke down the grammaticality contrast into adjacent (adjacent violation minus adjacent consistent) and non-adjacent (non-adjacent violation minus non-adjacent consistent) grammaticality contrasts, each of which revealed a single significant finding across the analysed regions of interest. The significant result under the adjacent grammaticality contrast is shown on the middle row of Figure 3.12. Here, the ERSP contrast shows that significant relative suppression was seen in the right hippocampus under the violation condition versus the consistent condition at 2400-3000 ms (at around 95-130 Hz only; note that, as this does not span the entire high gamma band, no equivalent period of significance is identified under the high gamma power analysis in the middle column, although the difference between the conditions is still weakly visible on this plot). Finally, under the non-adjacent grammaticality contrast (Figure 3.12, bottom row), no clear effect is visible under the ERSP contrast (rightmost column), but the mean high gamma power plot (middle column) instead here shows a brief period of significant relative high gamma *enhancement* in the right hippocampus under the violation condition versus the consistent condition at 1600-1800 ms. It should be noted that even if such responses were potentially also observed in the left hemisphere, coverage in the left hippocampus was insufficient to reach α (see Appendix 2: Supplementary figures, Figure 0.7 to Figure 0.9).

In summary, the conservatively computed group-level time-frequency contrasts, as in the time-domain analysis, revealed a limited number of effects overall. However, significant effects were

identified in the gamma band following the grammaticality manipulation in both right MTG and the hippocampus.

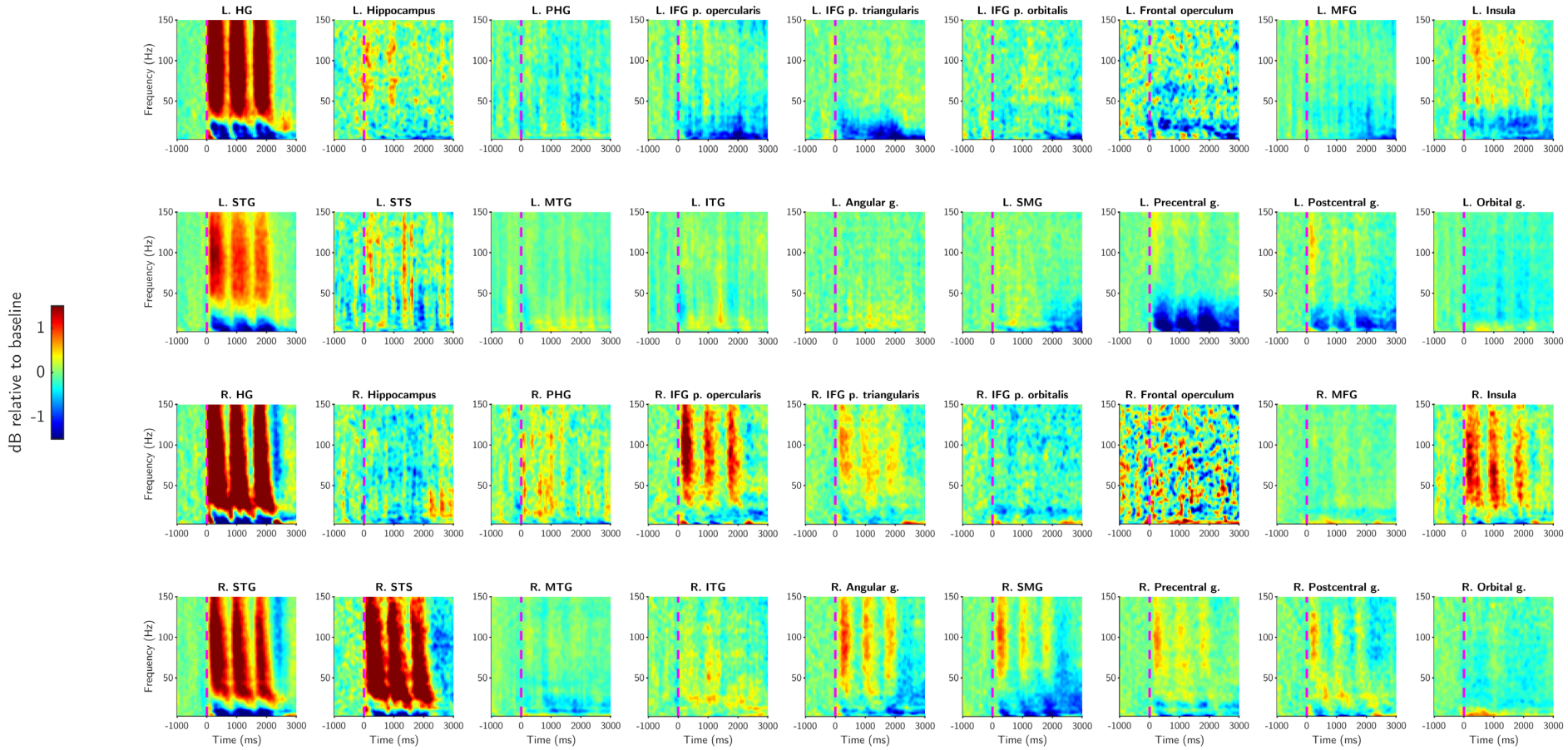


Figure 3.10: **Overall group mean (n = 12) event-related spectral perturbations (ERSPs).** Time is on the abscissa (-1000–3000 ms), and frequency on the ordinate axis (3–150 Hz), with colour denoting baseline-relative ERSP amplitude in decibels. Stimulus-driven responses aligned with the presentation of each syllable-word are visible in bilateral: Heschl's gyrus (HG), superior temporal gyrus (STG), insular cortex, pre- and post-central gyrus; left: hippocampus; right: superior temporal sulcus (STS), IFG p. opercularis, IFG p. triangularis, middle and inferior temporal gyrus, angular gyrus and supramarginal gyrus. In a number of cases (notably bilateral: HG and STG; right: STS, insula, IFG p. opercularis and triangularis, angular gyrus and SMG) this can be characterised as relative enhancement of β and γ band activity, and relative suppression of δ, θ and α activity. Less overtly stimulus-driven responses, still distinct from baseline, can be seen in left IFG as low-frequency suppression, especially in the pars triangularis.

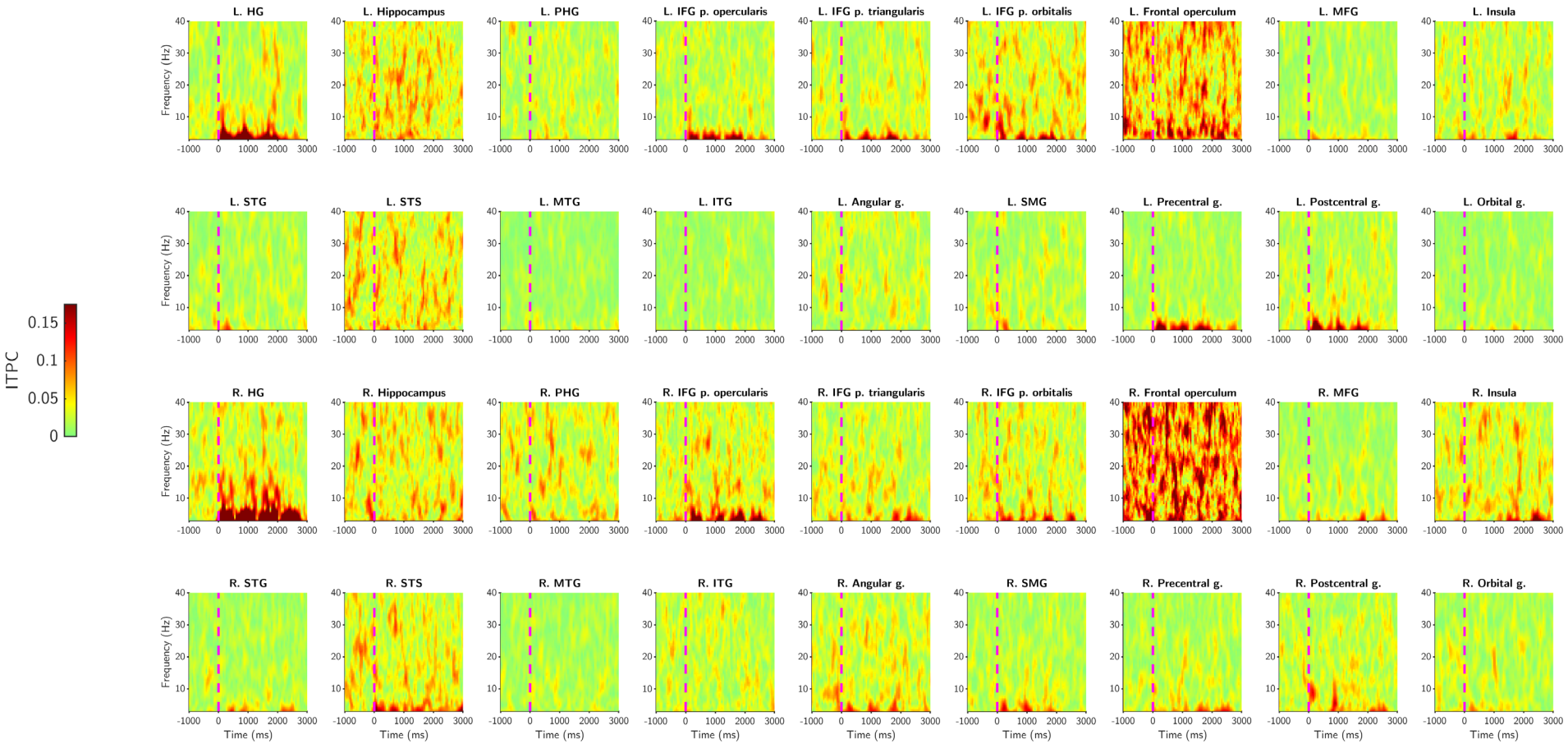


Figure 3.11: **Overall group mean (n = 12) inter-trial phase coherence (ITC) for analysed regions of interest (ROIs).** Time is on the abscissa (-1000–3000 ms), and frequency on the ordinate axis (3–150 Hz), with colour denoting ITC. Across all trials, phase-locked responses are observed at low-frequencies in a number of regions across the language network at times consistent with presentation of the stimulus items.

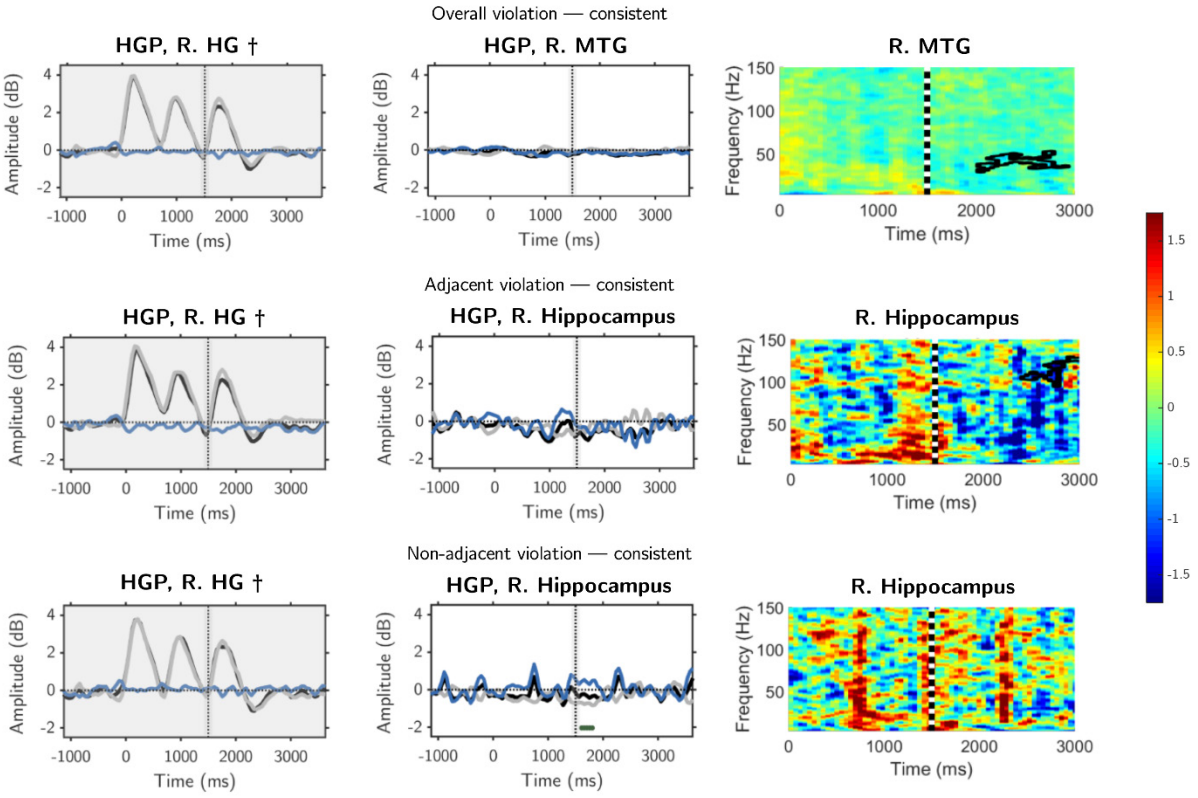


Figure 3.12: **A summary of selected cohort (n = 12) ERSP and high gamma power grammaticality contrasts.** For each contrast (overall violation – consistent; adjacent violation – consistent; non-adjacent violation – consistent) these figures include the only significant results found across the cohort. Vertical lines signify onset of the grammatical manipulation (3rd item). ***Left column:*** high gamma power (HGP, here always across 50 to 150 Hz) in right Heschl's gyrus, shown on the same scale as the middle column purely for comparison. These plots are grey and marked with a dagger (†) to indicate that coverage was insufficient to ever reach α using group-level permutation testing. ***Middle column:*** HGP in the same regions shown in the ERSP maps of the right-hand column. In both left and middle columns, black traces show mean "violation" responses, grey traces show mean "consistent" responses, and blue traces show the difference. Green dots show periods of significance after cluster-corrected permutation testing (α = .05, $\ll$10 000 permutations). ***Right column:*** ERSP differences are shown as colours on a scale from -1.75 to 1.75 dB. Black contours surround areas of significance as determined by cluster-corrected permutation testing (α = .05, $\ll$10 000 permutations). In the overall contrast, very subtle relative suppression is observed in right MTG between 2000 and 3000 ms at around 30-50 Hz (low gamma). However, in both the adjacent and non-adjacent contrasts, there are signs of brief high-gamma differences in right hippocampus under the two grammaticality conditions, although these meet significance under the ERSP and HGP analysis, respectively. In the adjacent "violation minus consistent" ERSP contrast, significant relative suppression can be seen in response to violations at 2400-3000 ms (at around 95-130 Hz only). Meanwhile, in the non-adjacent grammaticality contrast, HGP shows a brief period of significant relative high gamma enhancement at 1600-1800 ms.

3.5 Discussion

In this study, we assessed the ability of adult neurosurgical patients to learn regularities in a novel artificial grammar containing both adjacent and non-adjacent dependencies, and characterised intracranial responses to the sequences in the time and time-frequency domains. Behavioural analysis demonstrated that this task was extremely challenging for the patient cohort. Electrophysiological analysis, however, suggested engagement of key frontotemporal regions during the task, and revealed evidence of significant differences in the time- and time-frequency domains between neural responses under different syntactic conditions. A number of observed violation responses were consistent with canonical responses reported in the ERP literature, although typically these effects did not meet significance at the group level. Closer analysis suggested that grammatical violations modulated responses in key frontotemporal regions that differed for adjacent (**xAB**) and non-adjacent (**AxB**) sequences. Specifically, left IFG pars opercularis exhibited a significant difference in its responses to adjacent grammatical and adjacent ungrammatical sequences, but no such effect for non-adjacent sequences. All-trial time-frequency analysis revealed stimulus-driven modulation of frequency-specific power in a number of relevant regions, many exhibiting low-frequency phase-alignment in response to each of the syllables. Group-level time-frequency contrasts showed signs of weak relative low-gamma power suppression in right middle temporal gyrus (right MTG) in response to violations, across all sequences (via analysis of ERSPs); relative high-gamma power suppression in response to violations in the right hippocampus for adjacent sequences (via analysis of ERSPs); and relative high-gamma enhancement in response to violations in the right hippocampus for non-adjacent sequences (via analysis of high-gamma-band specific power). We discuss these findings in more detail below.

3.5.1 Behavioural findings

Compared to a healthy cohort previously assessed on a largely identical task (see Chapter 2), the neurosurgical patient cohort performed poorly on the *AxB* AGL paradigm, failing at the group level to exceed chance overall performance on the grammaticality judgement task either before or after implantation. Unlike the healthy adult cohort, the patients also did not exhibit a bimodal performance distribution, with only one or two high performers. A sliding window behavioural analysis suggested that performance in some participants varied somewhat over time, but no inferences can be drawn from the isolated breaches of performance thresholds identified in this

analysis. Consequently, there exists clear behavioural evidence for above-chance performance in only two of the participants (430 and 442; see Table 3.3), and only one participant maintained this level of performance across both runs of the task (430).

Whilst data collected in the healthy cohort previously revealed reaction time effects in both chance and high performers, here, again, significant reaction time effects were not revealed. The size of the neurosurgical cohort was necessarily restricted compared to the healthy cohort, so it is possible that it lacked the power to reveal this particular effect. Based on the results of Chapter 2, which showed a significant negative correlation between performance and self-reported fatigue, it is also possible that post-implantation reaction times and performance were impacted by multifactorial environmental factors affecting the epileptic patient population, and hospitalised patients in general. These include the effects of sleep deprivation, of pharmacological interventions such as analgesia, and potentially other iatrogenic factors (Lande & Gragnani, 2015; Giorgi et al., 2013; Cumming, 1984). However, we did expect higher pre-implantation performance. It is possible that other factors such as patients' nervousness about their upcoming surgery, or presurgical pharmacological interventions, may have affected performance during these sessions.

Alternatively, it has been suggested that temporal lobe epilepsy patients may exhibit deficits in sequence learning tasks by virtue of their pathology (Henin et al., 2021). There is insufficient evidence to reject this hypothesis here, especially in light of low pre-implantation performance. However, subjects had previously undergone neuropsychiatric assessment (see 3.3.1), and were screened for particular cognitive deficits that might be detrimental to performance. Moreover, given that the task was shown to be cognitively demanding in a healthy cohort (as reported in 2.4.2), it is equally possible that non-clinical demographic factors contributed to the poor performance.

In summary, we posited that explicit grammaticality judgement performance was low due to multifactorial environmental and demographic influences, whilst reaction time effects were most likely eliminated by a combination of fatigue and post-surgical medication, slowing responses and contaminating the recorded reaction time measurements. However, even without behavioural confirmation, we considered it likely that syntactic manipulations produced meaningful neural responses. This assumption was based on prior evidence that automatic neural responses to syntactic violations can be observed even in the absence of conscious awareness (Batterink & Neville, 2013).

3.5.2 Time-domain intracranial responses

The *z*-scored all-trial ERP results (Figure 3.6) revealed that responses to the final syllable of the sequences deviated considerably from baseline in many regions bilaterally, including critical regions of the left-lateralised frontotemporal language network such as left inferior frontal gyrus (IFG), although the coverage in many regions was insufficient to demonstrate significance under a group-level permutation analysis (see 3.3.5). Previously reported AGL tasks, however, have demonstrated similar task-related engagement of the fronto-temporal network, including prominent activation of IFG (see Batterink et al., 2019; Wilson et al., 2017).

The ERP contrasts revealed no evidence of significant responsiveness to sequence grammaticality across all sequence types combined (i.e. across both **xAB** and **xAxB**), but evidence of significant responsiveness to adjacent (**xAB**) sequence grammaticality in one region at the group level: left IFG pars opercularis. This is consistent with accounts that the processing of even simple adjacent dependencies engages IFG (Uddén & Bahlmann, 2012) and that stimuli need not be linguistic in nature to elicit expectation-violation responses in this region (Batterink et al., 2019). No significant responses were detected under the non-adjacent grammaticality contrast, though it may be that non-adjacent sequences preferentially engaged IFG subregions other than IFG pars opercularis, which could not have been detected under this analysis since no other IFG subregion had sufficient coverage to reach α under the time-domain permutation analyses (see Table 3.2 and section 3.3.5).

Although effects in only IFG pars opercularis met significance under the time-domain permutation analyses, a number of other regional ERP contrasts also featured characteristic non-significant positive and negative deflections reminiscent of left anterior negativity (LAN), P600 and N100 responses at the cohort level. In contrast to studies suggesting the LAN is an epiphenomenon resulting from interactions with N400 responses (Gonda et al., 2020; Tanner, 2019; Tanner & Van Hell, 2014), we saw little evidence in the regional contrasts or cortical projections of anything resembling the N400 response, which is consistent with popular accounts of a largely syntactic LAN/P600 and largely semantic N400 (Friederici & Meyer, 2004), given the lack of semantically meaningful content in our stimuli.

Our findings were potentially compatible with prior studies seeking to localise contributors to the LAN and P600 components, although the paucity of statistically significant effects at the group-

level limits the utility of these observations. Accomplishing source localisation from an EEG signal requires the researcher to approximate a solution to the *inverse problem* of which neuronal current sources generated which field potentials. This problem is ill-posed and, without constraints, there exist potentially infinite solutions to "which region generated what" (Grech et al., 2008). Consequently, the principal neuronal generators of specific ERP components have yet to be conclusively established. Multiple studies have mooted IFG as the source of the LAN, which appears to be eliminated in patients with left anterior lesions (Friederici et al., 1999; ter Keurs et al., 1999). Comparison of fMRI and EEG results has also suggested a role for the IFG in the generation of the P600 (Brouwer & Hoeks, 2013; van de Meerendonk et al., 2011), among other (parietal) regions theorised to contribute to the response. Our results were not entirely inconsistent with this view, showing the aforementioned statistically significant LAN-like response emanating from IFG pars opercularis in response to adjacent sequences, and as well as subthreshold (i.e. not statistically significant) LAN- and P600-like responses appearing to emanate from other subregions of inferior frontal gyrus. However, using a cortical projection method (NeuralAct; see Kubanek & Schalk, 2015) we also identified a number of other frontotemporal and parietal regions that appeared to generate positivities and negativities of a higher amplitude. ECoG is itself simply an intracranial form of EEG, and so the inverse problem still applies to results under it, but its spatiotemporal resolution and signal-to-noise ratio are higher than that of EEG due to the proximity of its electrodes to the cortex, allowing for accurate reconstruction of neural sources within millimetres of each contact (Todaro et al., 2019). As such, there is merit to this method of cortical projection, which estimates activity in the immediate vicinity of each contact. Whilst our results therefore add to a body of evidence suggesting that IFG has a role in syntactic processing and, more specifically, as a contributor to the LAN, they were not unequivocal, suggesting strong responses in temporal and parietal cortex that might influence this component at the level of the scalp. Such complex source interactions likely contribute to ongoing debate as to the origins of the LAN/P600 complex. In the future, triangulation of evidence under multiple imaging modalities (including fMRI, single- and multi-unit recordings, MEG, scalp EEG and ECoG), is likely to be essential to localising the sources of components indexing syntactic processing.

As previously mentioned, as well as computing overall grammaticality contrasts, we also separately determined differences in responses to adjacent (**xAB**) and non-adjacent (**AxB**) sequencing

violations under our AGL task. In IFG pars opercularis, in response to adjacent sequences, an early long-lasting negativity (200-500 ms post-manipulation) was observed in the adjacent violation case, whilst a non-significant positivity could be seen in the non-adjacent case. It is possible to reconcile these findings with more general accounts that the LAN is preferentially generated in response to local violations (Molinaro, Barber, et al., 2011; Molinaro, Vespignani, et al., 2011), as we predicted from findings in prior studies (see 3.2). Conversely, there is some evidence that a sustained P600 is more likely to be associated with the resolution of longer-range dependencies, perhaps indexing working memory for complex syntactic structure (O'Rourke, 2013; Gouvea et al., 2010), although no such findings reached significance in our analysis. Nevertheless, a close look at the mean ERP differences for both types of sequence (**xAB** and **AxB**; Figure 3.8) suggested that the deflections might not be as qualitatively distinct as they first appeared under the separate contrasts, as there were signs of similar negative- and positive-going deflections at canonical times in each contrast. Again, however, this could not be statistically verified at the group-level. If it could be verified, this would suggest that both the LAN and P600 are contributors to the ERP in the adjacent and non-adjacent cases, but with different weightings. Whilst ERP "components" are commonly taken to mean any marked positive or negative peaks or troughs in the waveform relative to baseline, such an account would be consistent with a more nuanced view of language-relevant ERP components as latent contributors to the final ERP (Brouwer & Crocker, 2017), which sum together to produce the observed effects. Future ECoG AGL paradigms run in a larger cohort of participants, and containing a wider range of dependency distance manipulations (e.g. **AxxB**, **AxxxB**), could support stronger characterisation of these seemingly orthogonal components.

3.5.3 Time-frequency domain intracranial responses

As in the ERP analysis, all-trial time-frequency analysis suggested extensive recruitment of frontotemporal and parietal regions relevant to language. In particular, under an all-trial ERSP analysis, many regions revealed a characteristic high-frequency enhancement (gamma, ~30-150 Hz) and low-frequency suppressive effect (<~30 Hz) aligned with the presentation of each syllable. All-trial intertrial coherence revealed low-frequency phase alignment correspondent with the syllable presentation in a number of critical regions, including left Heschl's gyrus, precentral and postcentral gyri, and all subregions of IFG. It also revealed signs of entrainment or auditory steady-state effects (Manting et al., 2021) following each sequence in regions including Heschl's gyrus and IFG.

Group-level ERSP and high-gamma power (HGP) permutation analyses followed a similar protocol to the ERP analyses, and therefore suffered from the same limitations caused by limited regional coverage at the level of individual participants. However, despite these limitations, three significant time-frequency effects were seen, namely in right middle temporal gyrus (MTG; seen in the overall grammaticality contrast using ERSPs) and right hippocampus (adjacent grammaticality contrast using ERSPs; and non-adjacent grammaticality contrast using HGP). Notably, IFG pars opercularis did not reveal effects under the time-frequency analyses.

Of particular interest, the right hippocampus showed signs of relative high-gamma suppression in response to adjacent violations, and relative high-gamma enhancement in response to non-adjacent violations, where the non-adjacent response exhibited an extremely low latency, and the adjacent response a marked delay. Interestingly, time-frequency analysis has previously revealed hippocampal sensitivity to prediction violations, but in the slow-theta range (2.5–5 Hz), rather than the gamma band as shown here (Chen et al., 2013). However, our findings are not physiologically implausible, as the hippocampus is thought to instantiate a theta-gamma code (Lisman & Jensen, 2013). Furthermore, hippocampal differences in gamma-band power have previously been linked to distinct processes of information and encoding and retrieval using human intracranial data (Griffiths et al., 2019). However, interpreting such scant effects is challenging under this task, especially since we still possess a relatively limited understanding of the functional relevance of distinct gamma-band frequency ranges (Prystauka & Lewis, 2019). Nevertheless, activation of subcortical structures, namely the hippocampus and striatum, is increasingly implicated in the learning of sequential statistical structure (Batterink et al., 2019; Opitz & Friederici, 2003). The hippocampus in particular may play a key role in the associative learning of sequential relationships and the prediction of upcoming items (Schapiro et al., 2014; Covington et al., 2018; see also 1.3).

As described, the data also revealed a very weak relative suppressive effect in right middle temporal gyrus in the low-gamma band in response to violation sequences. The *left* middle temporal gyrus has long been related to language function, specifically comprehension (Friederici, 2002; Price, 2010), and even to the P600 syntactic response (Friederici, 2011), although the contributions of the right MTG are less clear. As a consequence, the mechanistic relevance of this effect is hard to determine, although a probabilistic tractography study previously reported a correlation between right MTG white matter tract volume and performance on a pitch-based artificial grammar learning

study (Loui et al., 2011). In line with this finding, a recent study has also reported that EEG source localisation implicates right MTG in rule-based learning processes (Takács et al., 2021).

In summary, ERSP and HGP contrasts revealed a small number of interesting effects, similar to the time-domain analysis, that suggested neural sensitivity to the grammaticality of the stimuli. However, individual differences in functional frequency ranges may have diluted the group-level effects (as discussed, for example, in Doppelmayr et al., 1998). Similarly, it is possible that individual contacts of an ROI might contribute differently in the frequency domain overall, providing more information about a stimulus when aggregated in a more meaningful way than simple averaging (as in, for example, Tsuchiya et al., 2008). This potentially motivates the use of multivariate techniques in future, which account for high-dimensional features of a dataset, unlike the mass univariate analyses conducted here (Diedrichsen & Kriegeskorte, 2017).

3.6 Conclusion

We monitored neurosurgical patients undertaking an AGL task, recording and analysing behavioural and electrophysiological (ECoG/depth electrode) data. Despite an absence of behavioural confirmation, we found evidence of responsiveness to grammaticality in key regions of a frontotemporal network previously implicated in language, including a significant LAN-like time-domain effect in the left inferior frontal gyrus. We also found non-significant signs of deflections reminiscent of canonical ERP components in the time-domain, including the N100 and LAN, localisable to other regions of this network. The presence of the significant IFG LAN-like effect appeared to be modulated by sequence adjacency, consistent with prior accounts of the LAN preferentially indexing local morphosyntactic violations. Significant time-frequency effects were scant but notably included high-gamma-power effects in the right hippocampus, in line with previous findings suggesting hippocampal involvement in sequence processing (see 1.3).

Overall, the ability to ascertain group-level statistical significance using group-level permutation testing was limited in some regions by the number of clinical participants featuring intracranial coverage in those areas, which in turn was restricted by clinical availability. However, future studies may overcome these limitations by reducing the cognitive demands of the task, making it feasible to collect data in a wider variety of patients; and by using more advanced multivariate techniques to

account for per-trial variability and high-dimensional features of the data. Overall, these results seem to provide novel insights into the location and syntactic relevance of the LAN under a non-linguistic task.

Chapter 4. Structured sequence processing and combinatorial binding: neurobiologically and computationally informed hypotheses

Parts of this chapter (sections 4.1 – 4.7 inclusive)
have been published as

Calmus, R., Wilson, B., Kikuchi, Y. & Petkov, C.I. (2019) Structured sequence processing and combinatorial binding: neurobiologically and computationally informed hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences.* **375 (1791)**, 20190304.

and presented as

Calmus, R., Wilson, B., Kikuchi, Y., Kocsis, Z., Kawasaki, H., Griffiths, T., Howard, M. & Petkov, C. (2019) A computational model of complex combinatorial binding: Neurobiological simulations and hypotheses. *Society for Neuroscience 2019*. Chigaco, IL. 697.09 / BB11.

All work presented in this chapter was primarily conducted by the author, except where explicitly stated. The author conceived of the ideas, implemented simulations in MATLAB, produced the figures, and undertook the majority of writing. The co-authors provided discussion and helped to write and edit the published paper.

4.1 Abstract

Understanding how the brain forms representations of structured information distributed in time is a challenging endeavour for the neuroscientific community, requiring computationally and neurobiologically informed approaches. The neural mechanisms for segmenting continuous streams of sensory input and establishing representations of dependencies remain largely unknown, as do the transformations and computations occurring between the brain regions involved in these aspects of sequence processing. We propose a blueprint for a neurobiologically informed and informing computational model of sequence processing (entitled: Vector-symbolic Sequencing of Binding INstantiating Dependencies or VS-BIND). This model is designed to support the transformation of serially-ordered elements in sensory sequences into structured representations of bound dependencies; readily operates on multiple timescales; and encodes or decodes sequences with respect to chunked items wherever dependencies occur in time. The model integrates established Vector Symbolic additive and conjunctive binding operators with neurobiologically plausible oscillatory dynamics, and is compatible with modern Spiking Neural Network simulation methods. We show that the model is capable of simulating previous findings under structured sequence processing tasks that engage fronto-temporal regions, specifying mechanistic roles for regions such as prefrontal areas 44/45 and the frontal operculum during interactions with sensory representations in temporal cortex. Finally, we are able to make predictions based on the configuration of the model alone that underscore the importance of serial position information, which requires input from time-sensitive cells, known to reside in the hippocampus and dorsolateral prefrontal cortex.

The published content is presented in 4.1-4.7 almost entirely without revision, and from section 4.8 onwards is extended by the author (4.8, “Unifying neurocomputational accounts: extending the serial order code”).

4.2 Introduction

Natural environments are richly structured in both space and time. Substantial progress has been made in understanding the neurobiological bases of learned relationships between spatially or temporally separated elements (Kraus et al., 2013; Moser et al., 2015; O'Keefe, 1976). Moreover, prior research has established the importance of serial order for the brain (Lashley, 1951) and *binding problems*, whereby distinct sensory events are combined for perception, decision and action (Feldman, 2013), have attracted considerable interest and empirical enquiry (Gray, 1999; Singer, 1999).

Establishing relationships, or *dependencies*, between elements over time allows us to extract the structure of the sensory world and to make predictions about future events. However, understanding how the brain binds complex information distributed in time, building temporally organised structures that represent multiple linked dependencies, remains a considerable challenge facing the neuroscientific community. This sort of cognitive structure building is challenging for the brain to achieve because complex input must be discretized in time, resultant discrete items chunked and stored in memory, dependencies identified and related items bound for perception (or other purposes), and representations of multiple dependencies maintained concurrently in memory to be further manipulated (Fitch & Martins, 2014; Dehaene et al., 2015).

Human language – written, spoken, or signed – is a salient example of the complexity of the binding problem, because it features syntactically organized dependencies between semantic units (Tettamanti & Perani, 2012). Yet the problem of building complex representations is also relevant to complex action sequences (Fitch & Martins, 2014; Bickerton & Szathmáry, 2009; Rouault & Koechlin, 2018), music (Fitch & Martins, 2014; Jeon, 2014; Koelsch & Siebel, 2005), mathematics (Dehaene et al., 2015, 2004) and cognition in general (Bickerton & Szathmáry, 2009; Jeon, 2014). Moreover, some of these systems are not unique to humans, since songbirds can construct complex vocalization sequences (Sainburg et al., 2019), an ability supported by a forebrain neural system (Yu & Margoliash, 1996), and correspondences have been established between humans and a number of species in processing adjacent and non-adjacent sequencing dependencies (Wilson et al., 2018; Wang et al., 2015; Wilson, Kikuchi, et al., 2015). Thus, advancing our understanding of how complex structure can be built from sequential input in computational and neural systems is important for

developing better machine and animal models to understand both general principles and species-specific aspects of combinatorial binding.

In this article, we propose a blueprint for a neurobiologically informed and informing computational model of sequence processing (entitled: **V**ector-symbolic **S**equencing of **B**inding **IN**stantiating **D**ependencies or **VS-BIND**). The VS-BIND approach integrates: 1) advances in modelling combinatorial binding within simulated neural systems using Vector Symbolic operations; 2) insights from neuroimaging and neurophysiological evidence in human and nonhuman primates on neural correlates of structured sequence processing and working memory; and 3) dynamic mechanisms for manipulating population codes that can be incorporated in modern Spiking Neural Networks (Bekolay et al., 2014; Eliasmith, 2013). The approach allows us to plausibly transform internal representations, rendering these into both mathematically idealised and neurally simulated site-specific activity unfolding over time. Building on these foundational mechanisms, we focus on modelling chunk encoding and the binding of sensory items to represent adjacent, non-adjacent and more complex (hierarchically) structured sequencing dependencies. Our key objectives here are to motivate the approach, ground it in the relevant literature, and use it to generate distinct mechanistic predictions, the form of which, as will be seen below, depends both on the specific binding operations used and their configuration.

It is important to note that any model using combinatorial operators can only be described as *classically compositional* if combinatorial representations precisely reflect the meaning of all constituents and the relations between them. The operators used here are not classically compositional, but nevertheless serve an important purpose in allowing us to generate falsifiable predictions of neural mechanisms and correlates of structured sequence processing ripe for neurobiological testing across the species. To assist in this process we also share MATLAB (The MathWorks, Inc.) code, including a demonstration (DOI: 10.5281/zenodo.3464607; Calmus, 2019). Even on the basis of its structure alone and initial modelling, the VS-BIND approach posits a number of intriguing predictions.

In summary, the aims of this chapter are to motivate a coherent computational perspective on sequence processing that unifies a diversity of site-specific hypotheses traditionally beyond the remit of any single model, and to use this perspective to constrain emergent predictions on brain function.

4.3 Foundations of descriptive and computational models of structured sequence processing

Language relies on semantic and syntactic knowledge, supported by the detection of dependencies between phonemes, morphemes, words and phrases in sentences. The language binding problem features the rapid detection of lexical symbols and the encoding of complex syntactic regularities at multiple scales and temporal granularities.

A large volume of neurobiological data implicates a fronto-temporal brain system in various aspects of language processing (Bickerton & Szathmáry, 2009; Friederici & Chomsky, 2017; Hagoort, 2013). Neurobiological signals associated with the chunking and parsing of speech with respect to phrase boundaries have been identified in these regions (Ding et al., 2016; Ghitza, 2011; Nelson et al., 2017). Moreover, the temporal structure of speech or language content at different timescales (phonemic, syllabic, word or phrasal) produce stimulus- or context-driven neural entrainment at the relevant oscillatory frequency bands (Barczak et al., 2018; Giraud & Poeppel, 2012).

Behaviourally, a number of sequencing processes are now known to be evolutionarily conserved, including entrainment to rhythmic sensory input (Barczak et al., 2018; Lakatos et al., 2008). There is also information from Artificial Grammar Learning paradigms, which are used to establish dependencies between otherwise arbitrary auditory or visual items in a sequence, either via statistical or rule learning (Friederici, Bahlmann, et al., 2006; Kikuchi et al., 2018). Humans and a number of nonhuman animals can learn dependencies between sensory items next to each other in a sequence (*adjacent dependencies*), as well as dependencies further separated in time and by intervening items (*non-adjacent dependencies*; reviewed in Wilson et al., 2018). The learning of hierarchically organized dependencies by nonhuman animals is, however, contentious and it remains to be seen whether this ability is uniquely human (Wilson et al., 2017; Jiang et al., 2018).

Comparative neuroimaging work has identified brain regions in human and monkey frontal and temporal cortex involved in processing sequencing dependencies (Wang et al., 2015; Wilson, Kikuchi, et al., 2015). This has led to descriptive models of the brain-bases of structured sequence processing and the relationship with, and distinctions from, neurobiological processes involved in language (Wilson et al., 2017; Friederici, 2004; Petkov & Wilson, 2012). For the purposes of this paper, we will focus on the Wilson et al. (2017) descriptive neurobiological model of human and

nonhuman primate structured sequence processing, shown in Figure 4.1. Relatively simple sequencing relationships, such as between items that occur next to each other in a sequence, can be learned by both humans and monkeys and are seen to engage corresponding brain regions, particularly the ventral frontal and opercular cortex (Figure 4.1**A**, vFOC, bottom row) (Wilson et al., 2017). For non-adjacent dependencies (which increase working memory demands: store an item in memory long enough to link to its matched pair), it is less clear whether the frontal operculum, other inferior frontal areas (such as areas 44/45), and/or the dorsolateral prefrontal cortex are more involved. More complex (including hierarchical) dependencies engage inferior frontal areas 44/45 in humans (Broca's area; Figure 4.1**A**, middle and top rows; Friederici, Fiebach, et al., 2006). The hippocampus has also been implicated in structured sequence processing (Opitz & Friederici, 2004; Schapiro et al., 2017) and implicit learning (Jablonowski et al., 2018), but its mechanistic role within these contexts remains incompletely understood.

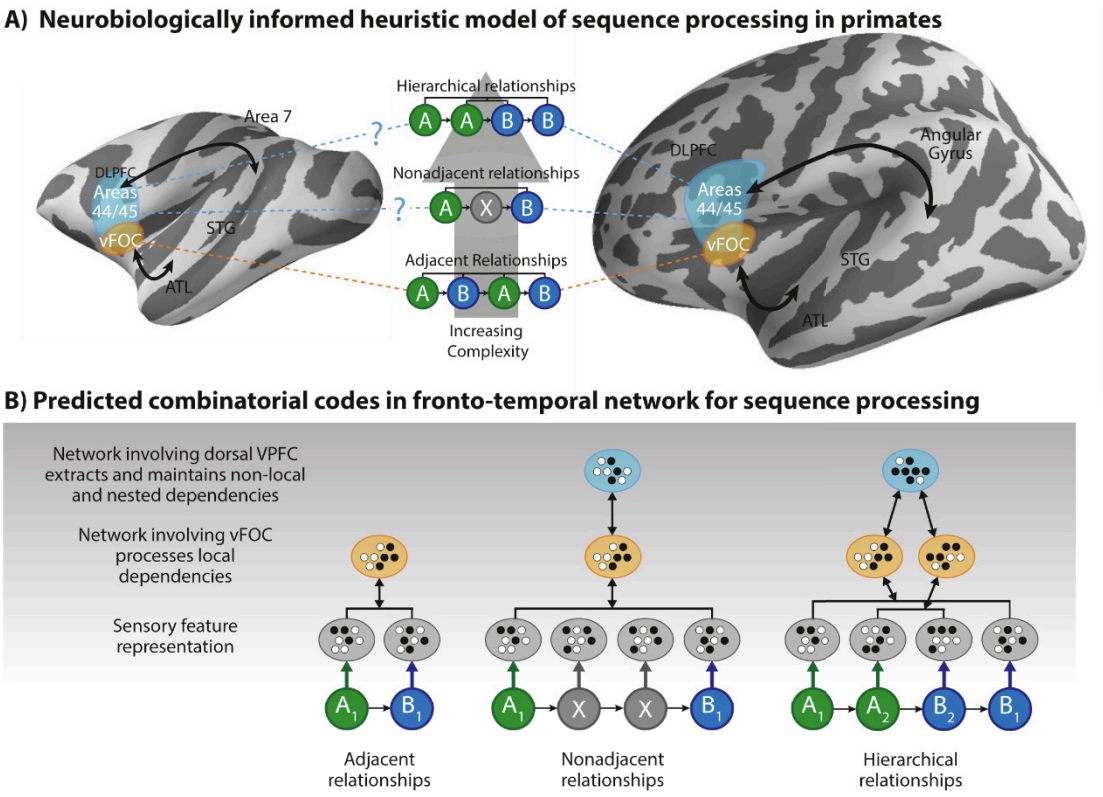


Figure 4.1: **Neurobiologically informed heuristic model of structured sequence processing, by Wilson, Marslen-Wilson & Petkov.** (A) Fronto-temporal regions involved in sequence processing, from Wilson et al., 2017 (Copyright © 2017 Benjamin Wilson, William D. Marslen-Wilson, and Christopher I. Petkov, CC BY 4.0). (B) Predicted combinatorial codes illustrated as neural patterns implemented by coordination between different regions.

Computational models of language or the processing of serial information provide compelling simulations of behavioural data (Cui et al., 2016; Gasser & Arbib, 2017). However, modelling underlying neural mechanisms presents additional challenges. David Marr's tri-level framework (Marr, 1982) famously defines three levels of description that are still widely applied in characterising any given model of the brain: the goals of the system (*computational* level), the cognitive processes required to reach this goal (*algorithmic* level), and the neural mechanisms required to instantiate them (*implementational* level). Poggio (2012) extended this framework by suggesting that models should also offer insights into learning processes and the evolutionary path that yields the system (see also Fitch, 2014). Although advances have been made in understanding many neuropsychological phenomena at individual levels of description, it remains desirable to advance understanding on multiple levels through holistic modelling approaches (Eliasmith & Kolbeck, 2015).

The fields of neuroscience and machine learning have been converging, in particular through the use of models incorporating functionally and anatomically distinct subpopulations of artificial neurons (Marblestone et al., 2016). Artificial Neural Networks (ANNs), including deep and recurrent neural networks (DNNs and RNNs), are the dominant connectionist modelling paradigm in use today, using iterative training procedures to tune synaptic weights between artificial neurons and establish network-level computations. RNNs are relevant to sequence learning since recurrent feedback allows them to integrate information over time (Hochreiter & Schmidhuber, 1997), whilst DNNs have revolutionised machine learning, increasingly inform neuroscientific analyses, and can generate neural correlates (Kar et al., 2019).

A likewise informative, but paradigmatically distinct approach is to model the brain at an algorithmic level, explaining behaviour and cognition in terms of computational processes that combine and transform cognitive symbols (see Samsonovich, 2010). Reconciling neural and cognitive perspectives is a longstanding challenge, but there exists a computational modelling subfield which has made considerable strides in this direction: *symbolic connectionism*. This approach seeks to produce ANN models with explicit support for combinatorial and symbolic operations (*neural-symbolic* networks). This is the subfield we look to in furthering our modelling aims, specifically using *Vector Symbolic Architectures* (VSAs; see below), which can be used alone, in ANNs, or in dynamic Spiking Neural Networks. Our use of VSAs has the benefit of generating

predictions on neural mechanisms of combinatorial binding throughout the fronto-temporal system involved in structured sequence processing.

4.4 Computationally modelling structured representations in neural systems: A brief overview of approaches

There exist a number of symbolic connectionist solutions, and many non-symbolic or non-connectionist cognitive architectures, each addressing various aspects of neural binding. Here, we briefly overview symbolic connectionist approaches before justifying our use of VSAs. For brevity, we restrict discussion here only to approaches that specify ways to build structured representations whilst explicitly supporting neurobiologically plausible implementation.

Three distinct methods of modelling binding predominate, although they can in many ways be viewed as complementary (Hummel et al., 2004). The first uses coordinated temporal synchrony (see LISA, in Hummel & Holyoak, 2003, 1997; and SHRUTI, in Shastri & Ajjanagadde, 1993) or asynchrony (see DORA, in Doumas et al., 2008) to unify constituents. The second uses uniform grids of integrating circuits to create a stable memory for bindings (Neural Blackboard Architectures, NBAs; van der Velde & de Kamps, 2006). The third, encompassing VSAs, principally uses conjunctive spatial coding of abstract vector representations to associate items, via tensor products (Smolensky, 1990; Halford et al., 1998; Wilson et al., 2001), circular convolution (Plate, 1995) or other defined transforms (Kanerva, 1994; Gosmann & Eliasmith, 2019).

Although commonly contrasted, spatial and temporal binding mechanisms are not mutually exclusive. Conjunctive spatial coding and temporal synchrony/asynchrony are known to be complementary and can be thought of as different perspectives on the same dynamic process (Hadley, 2007). Furthermore, conjunctive coding is considered appropriate for long-term storage in temporally coordinating models (Hummel et al., 2004; Martin & Doumas, 2017), whilst vector-based methods can operate within dynamic frameworks likewise subject to temporal influences (Eliasmith, 2013), which we will consider in more detail later. Thus, spatial and temporal binding approaches are not diametrically opposed but rather place different emphases and explanatory burdens on mutually informing aspects of neural coding. Although VSAs tend to be conceived of as static systems, we specifically advocate their use within a dynamic framework (for example Nengo,

Bekolay et al., 2014; Eliasmith, 2013) to incorporate advantages of both temporal and population coding.

The neural population codes employed by these models vary. In this regard, *localist* representations (which exhibit one-to-one or many-to-one mappings between features and neural activation) are typically contrasted with *distributed* representations (which exhibit many-to-many mappings). Distributed representations may be *dense* or *sparse*, which in the latter case means that a low proportion of neurons are active within a population at any one time. There is evidence that both localist (Roy, 2017; Bowers, 2009) and sparse distributed representations are utilised by the brain (Wixted et al., 2014; Willmore & King, 2009; Rolls & Treves, 2011), and there are advantages to both encoding strategies for neural systems. For instance, localist representations exhibit the lowest possible interference between encodings, whilst sparse distributed representations exhibit graceful degradation in performance in the presence of increasing noise (Werning et al., 2012). Sparse distributed vectors exist on a representational continuum that allows them to demonstrate characteristics of either localist or distributed encodings depending on their sparsity (Doumas & Hummel, 2005). This flexibility motivates the use of abstract vector systems that define combinatorial operators over sparse distributed representations. VSAs (Gayler, 2003), which we use to implement VS-BIND, accomplish this.

These models also differ in terms of how semantic representations are instantiated (Logie, 2007). Some define the semantic structure of relational encodings at the neural level, generating explicit *role-filler* (Hummel & Holyoak, 2003; Shastri & Ajjanagadde, 1993) or *symbol-argument* bindings (Halford et al., 1998). These models provide clear mechanisms to support compositional relational encodings of semantic knowledge, where the whole perfectly reflects its parts, a feature considered important for linguistic modelling (Hummel & Holyoak, 2003, 1997; Shastri & Ajjanagadde, 1993; Doumas et al., 2008). By comparison, VSAs do not inherently specify neural implementations at all, but are rather supported in this regard by broad theoretical frameworks that specify mappings between abstract vectors and neural population codes (for example, the Semantic Pointer Architecture, used in Nengo – see Eliasmith, 2013 and Bekolay et al., 2014 – or Integrated Connectionist/Symbolic Architecture, ICS; see Smolensky & Legendre, 2011).

Structured sequence processing typically focusses on ordering relationships and can refer to operations on meaningless items (e.g., nonsense words or abstract visual shapes). We consider cognitive architectures incorporating VSAs (Smolensky, 1990; Plate, 1995; Kanerva, 1994; Gosmann & Eliasmith, 2019) to have particular strengths appropriate to their use in a neurobiologically plausible model of structured sequence processing. Firstly, using existing tools, VSAs can act as a bridge to multiple modelling paradigms. For example, Nengo, a library for large-scale dynamic neural simulation (Bekolay et al., 2014), supports VSA encodings within spiking neural networks through the use of the Semantic Pointer Architecture. Bayesian computations (Sharma et al., 2017) and attractor dynamics (Gilra & Gerstner, 2017) have also been instantiated within this spiking neural network system. Secondly, VSAs are highly scalable solutions possessing substantial storage capacity for high-dimensional information (Crawford et al., 2016). Thirdly, as we describe shortly, VSAs permit the definition of relationships at the algorithmic level using an easily interpretable algebra. Finally, VSAs remain neurocomputationally plausible (Eliasmith, 2013) whilst being relatively straightforward to implement (Plate, 1995; Kanerva, 1994; Levy & Gayler, 2008). To describe our approach, we next outline VSA principles and operators, before explaining how temporal dynamics can control such operations. We conclude by describing the role of these mechanisms within a neurobiologically plausible model of structured sequence processing, VS-BIND.

4.5 Combinatorial population coding with Vector Symbolic Architectures

In VSA models, the basic units of representation are high-dimensional vectors. These typically sparse representations (see Figure 4.2, **A** and **B**) can be visualised directly (Figure 4.2, left panel) or encoded as neural activity using multidimensional tuning functions (Figure 4.2, right panel) (Bekolay et al., 2014; Eliasmith et al., 2004).

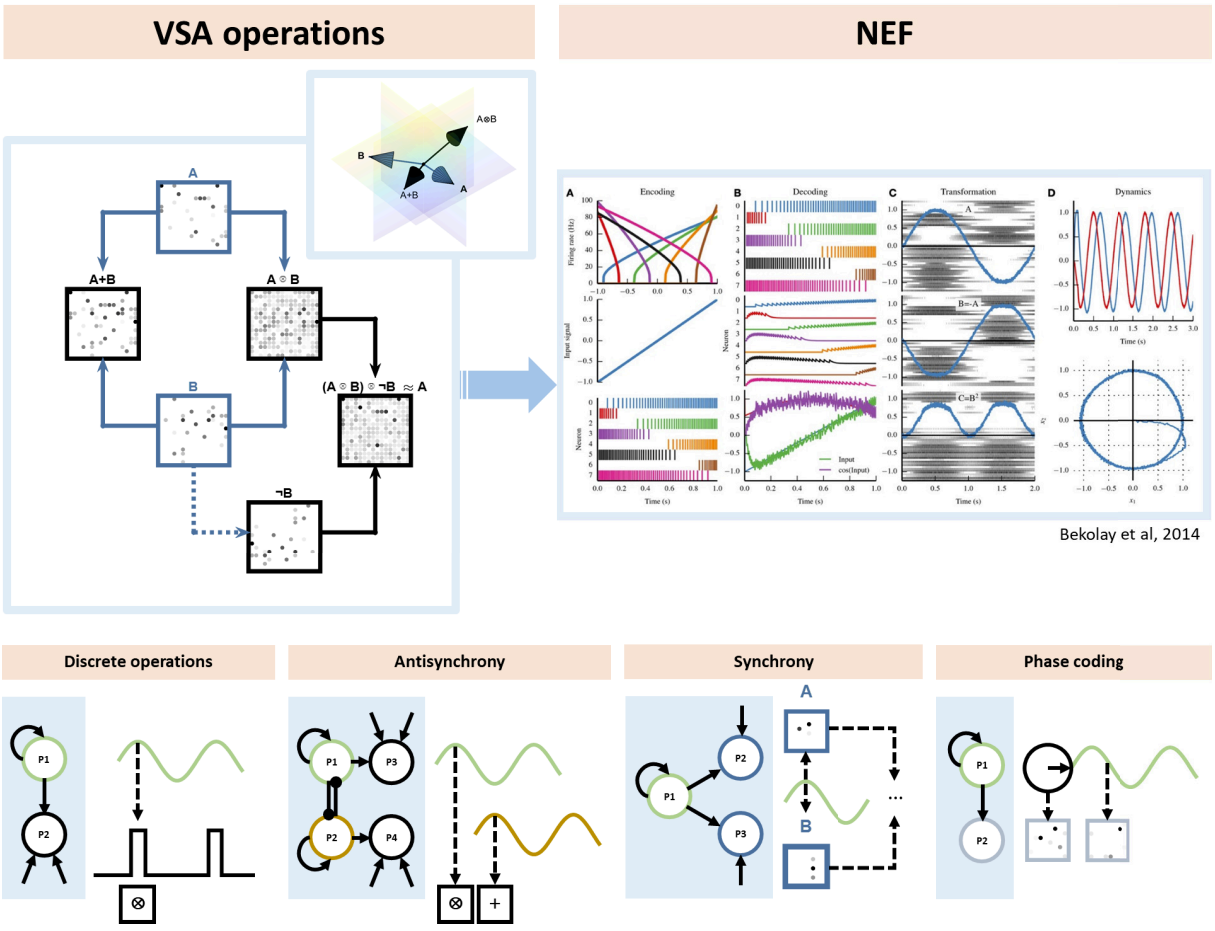


Figure 4.2: **Spatial and temporal coding within a spiking model**. ***Top row, left panel:*** Vector Symbolic Architecture (VSA) operations using circular convolution to accomplish binding ("A ⊗ B"; Plate, 1995). Sparse, high-dimensional random vectors represent distinct symbols **A** and **B** (blue text). To aid visual comparison, vectors are shown reshaped into squares, meaning the dimensionality of each vector equals the number of "pixels" in each box (here, 256-dimensional, plotted as 16 × 16). These vectors can also be considered *directions* in high-dimensional space (**inset**, projected down to 3-D using PCA). Results of the VSA operators are shown (**left main panel,** black text, clockwise from left: *superposition*, *binding*, *unbinding* and *involution*; see manuscript), with arrows indicating the flow of operands. A noisy recovered vector (bottom right square) can be cleaned up with an autoassociative memory to produce **A** (top square). ***Right panel*** (from Bekolay et al., 2014; Copyright © 2014 Bekolay, Bergstra, Hunsberger, DeWolf, Stewart, Rasmussen, Choo, Voelker and Eliasmith, CC BY 3.0): Core properties of the Neural Engineering Framework (NEF; Eliasmith, 2013) as implemented in Nengo (Bekolay et al., 2014). To the left are tuning curves of individual neurons (**A, top plot**). In vector terms, each neuron fires maximally to its own *preferred direction*. Nonlinear *encoding* of an input signal (**A, middle plot**) yields spike trains for each neuron (**A, bottom plot**). *Decoding* (**B**) is possible using linear methods. Combining decoding with encoding, one can determine synaptic weights representing transformations between populations (**C**). Here, VSA representations are simply high-dimensional signals encoded like any other. Operations like convolution can be learned by simulated spiking networks incorporating Spike-Timing-Dependent

Plasticity (STDP) in Nengo (Eliasmith, 2013). Finally, dynamic signals can be represented (**D**), of relevance for understanding oscillatory mechanisms (Ding et al., 2016; Nelson et al., 2017; Barczak et al., 2018). Nengo is agnostic about neural models, with many spiking models available (Hodgkin & Huxley, 1952; Izhikevich, 2003; Stewart et al., 2009). ***Bottom row:*** Since the NEF provides mechanisms for spatial (VSA operations) and temporal (dynamic) manipulation of representations, possibilities exceed that of a static system. Simple interactions between segregated populations (*networks shown in blue boxes*) lead to controlled functional relationships. Thresholded dynamic activity, e.g. arising from an oscillator (P1, ***leftmost panel***) can trigger discrete combinatorial operations. These operations can be segregated over time (***middle left panel***) through control by interacting, antisynchronous oscillations (P1 and P2). Likewise, common driving signals can synchronously strengthen representations in disparate regions (***middle right panel***, P1 multiplicatively modulating P2 and P3) for downstream processing such as feature binding. Finally, the *phase* of an oscillator, rather than its amplitude (***rightmost panel***) can drive downstream encodings such as those of relative position.

Symbolic vectors can be recombined using specific, reversible combinatorial operators to create new representations containing information on their constituents and the relations between them (Smolensky, 1990; Plate, 1995; Kanerva, 1994). Inputs to these operations can either be atomic vectors (those that are not compound representations) or the results of previous operations. Atomic vectors are often randomly generated in VSA models, but can also be generated by compressing even higher dimensional input. Using this latter approach, semantically similar concepts cluster together in vector space (Eliasmith, 2013).

The power of VSA approaches lies in their combinatorial operators. Although these are broadly common to all VSAs, here we utilise only one VSA, that of Plate (1995). This VSA uses the following operators: *superposition* (or *bundling*; point-wise vector addition of inputs), *binding* (a conjunctive operator) and its inverse, *unbinding* (also termed *release*, which here relies on *inversion*, equivalent to logical "not"), and a vector *comparison* operator for readout of results. Note that within VSA terminology, only the conjunctive operation is known as *binding*, but both this specific operator and superposition fit the wider definition of binding in the broad context of neural binding problems (Feldman, 2013). The more specific VSA nomenclature, which we adhere to hereafter, is helpful because it imposes constraints on the potential neural mechanisms involved in each of these combinatorial processes.

Using the full set of VSA operators, algorithmic manipulations are undertaken easily. The operators are best demonstrated with just two inputs (here **A** and **B**; typical VSA operators are illustrated in the left panel of Figure 4.2). **A** + **B** (*superposition*) yields a vector *correlated* with both A and B. This is simple vector addition, essentially overlaying the sparse inputs. **A** ⊗ **B** (*binding*) is a conjunctive operator that yields a vector approximately *orthogonal* to both A and B, and as such is poorly correlated with either input. In this VSA, binding is calculated through circular convolution (Plate, 1995), which is just one possible way to create an output vector of the same length as one input alone (see also Kanerva, 1994; Gosmann & Eliasmith, 2019). This feat is possible because the operator encodes a *reduced representation* of both inputs, which can be unbound as described below.

Reduced representations are an important feature of a number of VSAs. Without them, conjunctive operations on input vectors of length N each result in output vectors of length N^2 (leading to exponential increases in vector size, a *combinatorial explosion*). This is a characteristic of

the ancestral VSA, Smolensky's *tensor product binding* model (Smolensky, 1990; Smolensky & Legendre, 2011). Plate's VSA (1995), by contrast, overcomes this scaling problem by making use of compressive representations known as *Holographic Reduced Representations* (HRRs). Reduced representations allow a single model to support repeated (or recursive) operations on vectors without dimensionality increasing; the limiting factor in using reduced representations is instead that the process is lossy, so repeat operations cumulatively degrade the output vector. It is partly for this reason that this approach cannot be considered classically symbolic or perfectly compositional. However, lossy encoding recapitulates natural limits on working memory and the depth of recursively nested structures that can be constructed or comprehended in natural language (de Vries et al., 2011).

Here*, unbinding* is in essence binding, but with a change to one of the operands. Values may be unbound or *released* from a bound representation by computing a new binding between it and the *approximate inverse* of one of the original inputs with respect to the binding operator (from hereon in, just named the *inverse*, "¬"; see **¬B,** Figure 4.2). Inversion is accomplished by simply permuting all but the first dimension of **B**, which at a neural level means rerouting input dimensions using a distinct pattern of synaptic weights. It is easier to understand the inversion operator by demonstrating its use in unbinding; the vector symbolic formula **(A ⊗ B) ⊗ ¬B** shown in Figure 4.2 depicts this process. Here, we bind the bracketed representation with a "key" containing the inverse of the element we know to be linked to the vector we wish to recover. Without such a key it is impossible to retrieve the contents of a binding via circular convolution. This means that VSA operations must be configured carefully to ensure that bound information can be retrieved in a plausible way.

To interrogate the result of unbinding or any VSA operation, we can use the comparison operator to assess the similarity of the output to a defined vocabulary of representations. This can be useful as part of an autoassociative memory, or to control downstream actions or operations. For example, let **R** = **(A ⊗ B) ⊗ ¬B**. Here, undertaking a comparison between **R** and a set of two vectors {**A**, **B**} would reveal that **R** is highly similar to **A** and highly dissimilar to **B**. Comparisons can be conducted on real-valued vectors using correlation or the vector dot product (where a smaller angle between input vectors indicates stronger similarity). The dot product in particular could be

implemented by any neuron; it is equivalent to simply summing all presynaptic potentials weighted by their respective synaptic weights over a short period of time (Rolls & Treves, 2011).

Use of VSA operations is by no means restricted to elementary computations. Mathematically, the superposition ("+") and binding ("⊗") operators are designed to exhibit associativity, commutativity and distributivity akin to their scalar analogues, addition and multiplication (Plate, 1995). For example, if **A**, **B** and **C** each represent vectors, the vector symbolic formula $\mathbf{A} \otimes (\mathbf{B} + \mathbf{C})$ is equivalent to $(\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{C})$, just as one would expect algebraically. This makes the results of multiple VSA operations predictable and transparently interpretable. In a similar fashion, unbinding still functions if we superpose multiple bindings, as in $\mathbf{R} = (\mathbf{A} \otimes \mathbf{B}) + (\mathbf{C} \otimes \mathbf{D})$. In this case, $\mathbf{R} \otimes \neg\mathbf{D}$ will result in the value ($\mathbf{C}$ + *noise*), recovering a representation highly similar to **C**.

Are such operations neurobiologically plausible? Plausibility of the neuronal arithmetic at a basic level is well supported; additive and multiplicative functions, which are sufficient to compute all of the VSA operations described here, including circular convolution, are established neural processes (Silver, 2010). Multiplicative and divisive functions abound in the cognitive neurobiological literature, as feedback influences on neural responses (Kikuchi et al., 2019; Reynolds & Heeger, 2009). The circular convolution operator could also be substituted for a number of alternative conjunctive distributed operators, for example vector-derived transformation binding (Gosmann & Eliasmith, 2019). There is a good deal of evidence consistent with the presence of conjunctive distributed encodings in associative areas such as retrosplenial cortex (Beyeler et al., 2019), the ventral visual stream (Erez et al., 2016), and hippocampal CA1 and CA3 subregions (Komorowski et al., 2009).

How do we distinguish superposition from multiplicative (Doumas & Hummel, 2005) operations like convolutional binding in high dimensions? We expect that neurons instantiating additive operations should demonstrate linear responses to linear combinations of features of constituent representations, whilst multiplicative operations should result in nonlinearly selective responses. Empirical evidence already suggests that superposition cannot be the only combinatorial function, since linear and nonlinear mixed selectivity are *both* prevalent within the lateral prefrontal cortex and elsewhere (Rigotti et al., 2013; Parthasarathy et al., 2017). Thus, there is evidence for the

broad classes of functions to which the VSA operators belong in regions supporting cognitive function.

Simulations likewise demonstrate plausibility; Plate's VSA operators, discussed here and shown in Figure 4.2 (Plate, 1995), form the combinatorial backbone of Eliasmith's Semantic Pointer Architecture (SPA), a neurobiologically plausible representational framework underlying Nengo, a Python library supporting the construction of spiking neural models (Bekolay et al., 2014; Eliasmith, 2013). The SPA proposes that high dimensional representational vectors arising in sensory cortex are compressed by similarity-preserving dimensionality reduction, and that the brain undertakes vector symbolic operations on these reduced representations. Nengo supports the transcoding of these cognitive representations into spiking neural representations by applying a set of fundamental neural encoding and decoding principles, the Neural Engineering Framework (NEF; Bekolay et al., 2014), summarised with reference to the original paper in Figure 4.2 (right panels and figure legend). Under this system, each spiking neuron contributes to a distributed encoding of an underlying latent vector representation; that is, there is a many-to-many mapping between the dimensions of a cognitive vector (itself already a distributed representation) and individual neurons. The resulting spiking activity is inherently dynamic, and thus suitable for manipulation over time by temporal mechanisms within the same dynamic framework.

4.6 Dynamically coordinating combinatorial operations with temporal mechanisms

As we have seen, population coding and temporal mechanisms are both functionally important to any account of domain-general structure building (Ding et al., 2016; Doumas & Hummel, 2005). Here, we introduce basic oscillatory principles that serve as temporal mechanisms within VS-BIND.

Neural oscillations in the brain reflect temporally coordinated responses of neural populations (Whittington et al., 2011; Buzsáki & Wang, 2012). Oscillatory signals can also result in oscillatory *coupling* across frequency bands that reflect coordinated interactions within and between brain regions (Buzsáki & Wang, 2012; Fries, 2015). Theta-gamma coupling, for example, is associated with cognitive function and is also seen during structured sequence processing tasks (Buzsáki & Wang, 2012; Kikuchi et al., 2017).

We can instantiate dynamic relationships, including coupling, within a spiking neural model such that they undertake functionally useful coordinating or multiplexing roles in neurobiologically plausible ways (Figure 4.2, **bottom row**). For example, a self-connected population (**leftmost panel**, P1) can generate oscillatory dynamics that, when thresholded, serve as the trigger for discrete vector symbolic operations on multiple inputs (P2). Likewise, interacting oscillators (P1 and P2, **middle left**) can exhibit antisynchrony, segregating discrete operations such that they do not interfere. Temporal synchrony can also be simulated (**middle right**) through top down influences on multiple neural populations, for example by controlling gain, such that downstream operators (for example, VSA operators) act on coordinated inputs. Here, temporal coordination serves to functionally associate two vector representations that would otherwise remain separate.

The dynamic context in which these spatial operators act is crucial, because it serves to mitigate the concern that, due to their multiplicative interactions, such spatial bindings are too variable to support generalisation over classes of their inputs, and thus insufficient as relational encodings, a problem characterised in the literature as *violation of role-filler independence* (Doumas & Hummel, 2005). The explicit segregation of representations in space (Figure 4.2, **A** versus **B**) and time (Figure 4.2, top left panel, upstream/downstream) means that multiple neural ensembles concurrently instantiate different components of a combinatorial representation. Variability in the downstream binding of **A** and **B** would not prevent either of these two populations from generalising over their inputs. Constituents may be dynamically bound or unbound as needed to segregate or aggregate information.

Finally, oscillatory signals can unidirectionally coordinate activity elsewhere, for example on the basis of phase (Figure 4.2, **rightmost panel**), thus potentially exhibiting phase-amplitude coupling effects. This produces behaviour consistent with conceptual models on the role of theta-gamma coupling (Lisman & Jensen, 2013) observed in the brain (Buzsáki & Wang, 2012; Canolty et al., 2006). VS-BIND exhibits phase-amplitude coupling as an explicit functional property of relative position coding.

The subject of dynamic coordination is returned to in more depth in sections 4.7 ("Network-level mechanistic hypotheses derived from VS-BIND") and 4.8 ("Unifying neurocomputational accounts: expanding the serial order encoding").

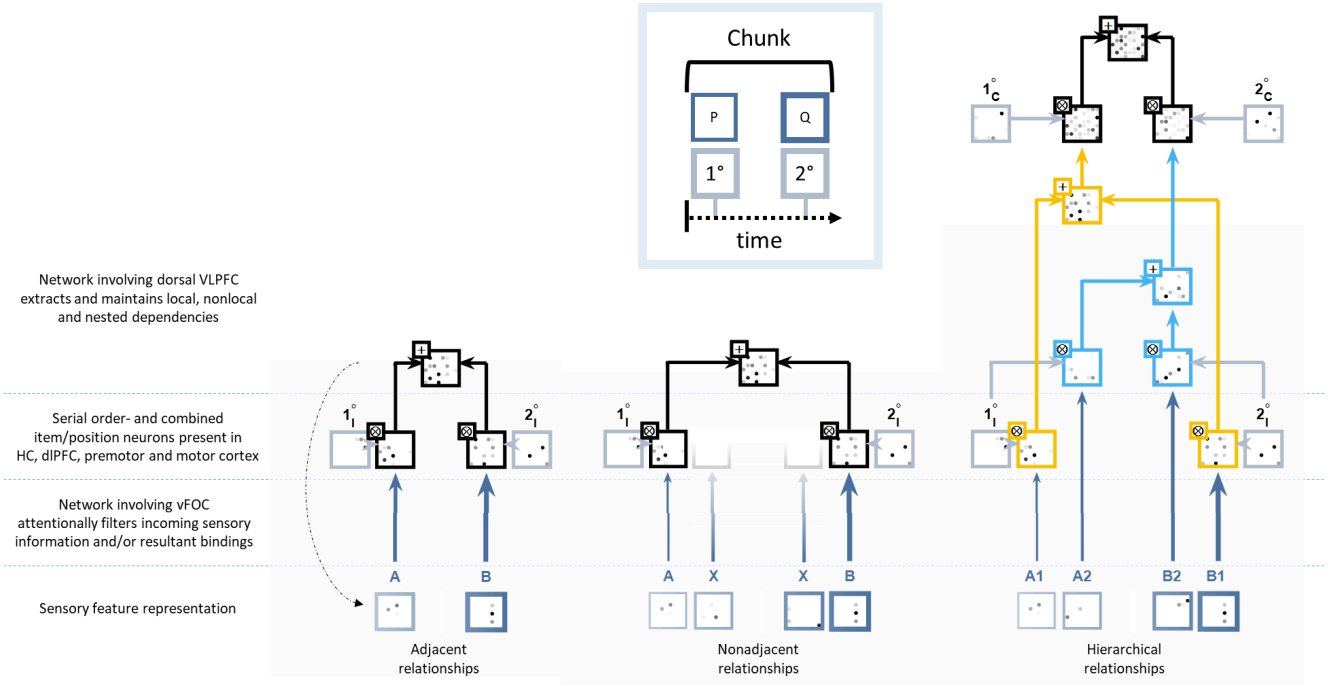


Figure 4.3: **Neurobiologically informed vector symbolic encoding of sequence structure.** Vector symbolic operators can account for the processing of a variety of sequencing dependencies. The solid arrows in these charts indicate the flow of information during *encoding* of a stimulus sequence only. These describe transformations of latent vector symbolic representations, as opposed to neural activation patterns. Representation strength is denoted by the shading and thickness of each box border. For clarity, representations are shown separated along the horizontal axis, though separate boxes do not necessarily imply separate neural populations are engaged, especially if describing identical computations, which could be undertaken by neurons of a single region. We suggest sensory representations (**bottom row**) are maintained within SMA (not shown) and retrieved as needed. Operations unfold dynamically following principles outlined in Figure 2. The final encoded sequence representation is found at the top of each diagram. Each is a reduced representation whose constituents can be inspected without serially unpacking all bindings; the superposed final result of the *adjacent relationship* encoding (**leftmost diagram**), for example, can be interrogated to recover its secondary element by simply binding it with **¬2°I**. Serial elements packaged into a single representation are considered to be *chunked* in the traditional sense (**inset box**), but identical

operations can be applied to nonadjacent (**middle diagram**) or nested pairs of elements (**rightmost diagram**), using separate item (**I**) and chunk (**C**) position encodings. Selective fading along the vertical axis represents salience filtering in the nonadjacent example (**middle**). Finally, the dashed, curved arrow shows just one case in which *sub-symbolic* feedback from a cognitively abstract cortical region might ultimately influence the representation of individual elements in sensory cortex (there may be many such pathways, but one exemplar is shown). Thus, although the figure, for simplicity, suggests VS-BIND is largely a feed-forward model, feedback influences feature and can, for example, allow certain areas to influence sensory cortical representations.

4.7 Network-level mechanistic hypotheses derived from VS-BIND

Our objective within the remainder of this paper is to outline the specific combinations of operations required to support dependency encoding during sequence processing, generating neurobiological hypotheses for adjacent, nonadjacent and hierarchical dependency encoding. The MATLAB demo provided at DOI:10.5281/zenodo.3464607 further illustrates key principles outlined here.

4.7.1 Adjacent dependencies

We first consider the binding of two items following each other closely in time. At the vector symbolic level, ordered sequences can be unambiguously represented using one of two principal methods. The first is by encoding elements with respect to each other (*chaining*). However, behavioural findings in multiple species do not convincingly support this approach (Choo & Eliasmith, 2010). An alternative method, encoding each element with respect to a serial *positional tag* (*Sequence* = 1^{st} * *Item1* + 2^{nd} * *Item2*, and so on), has found greater support. Behavioural results in songbirds suggest a reliance on positional cues during sequence recognition (Comins & Gentner, 2010). Likewise, during recall, humans and non-human primates are more likely to confuse items in different memorised sequences if those items share the same ordinal position across sequences (Dehaene et al., 2015). Positional influences on recall provide behavioural evidence that ordinality is incorporated into the encoding of sequences even when position is not explicitly featured in a task, offering support for positional tagging.

Detection of ordinal serial position, an essential component of sequence encodings within VS-BIND, we posit involves dorsolateral prefrontal cortex (DLPFC), motor and premotor cortex. Electrophysiological findings in non-human primates align with this account, revealing populations of cells in each of these regions involved in consistently encoding serial position irrespective of stimulus identity (Carpenter et al., 2018; Petrides, 1991). Studies of the hippocampus likewise reveal temporal coding relative to stimulus presentation, hypothesised to form part of the context for later retrieval (Long & Kahana, 2019). There is thus a neurobiological basis for the crucial role played by explicit positional tags within our model (Figure 4.3, **light blue-grey boxes**). Within VS-BIND, positional tags are considered to be deterministically but flexibly generated by the brain.

Here, tags follow the nomenclature "primary" and "secondary" ("1°" and "2°") rather than the ordinal absolutes "first" or "second", in line with human and nonhuman primate behavioural (Henson, 1999; Endress et al., 2010, 2009) and electrophysiological (Carpenter et al., 2018; Long & Kahana, 2019) evidence suggesting that, within sequences of words or actions, ordinal position is encoded relative to sequence boundaries rather than in absolute terms. Within our model, positional tags are anchored to the boundaries of perceptual chunks (Figure 4.3, **box, inset**), by decoding stimulus-entrained oscillatory phase (see Figure 4.2) such that both positional tags and sequence items are derived from the stimulus. It is important that each tag remains orthogonal to the last, which is a requirement for later unambiguous recovery of specific elements in a sequence.

To encode a sequence, continuous input over time is first discretized. These discrete sensory items are bound to distinct positional tags (to form position-item representations) and superposed in a decaying, recurrently connected working memory buffer. This *sequence buffer*, likely supported by SMA and pre-SMA (Cona & Semenza, 2017), therefore maintains a linear, ordered representation of the input sequence (i.e. $Sequence = 1° \otimes Item1 + 2° \otimes Item2 + 3° \otimes Item3$). From this, individual sensory representations can subsequently be retrieved (via $Item \approx Sequence \otimes \neg Position$) and recoded to reflect dependencies between items or chunks (Figure 4.3, **bottom row,** showing retrieved items). During this recoding process, maintained representations (**A**, **B**, or irrelevant intervening items, **X**) serially accumulate within a *dependency buffer* over time (i.e., moving rightwards) where they can be used in increasingly complex binding operations (moving upwards). The SMA encoding steps are comparable to the VSA approach in the Ordinal Serial Encoding (OSE) working memory model of Choo & Eliasmith (2010), which is capable of modelling behavioural characteristics of serial recall such as working memory *primacy* and *recency* effects. However, unlike the OSE model, VS-BIND incorporates centrally coordinating oscillatory activity and uses boundary-relative (rather than absolute) ordinal codes. Moreover, subsequent to serial encoding within working memory, VS-BIND describes the encoding of various dependencies, which need not be linear in organisation (Figure 4.3).

To encode dependencies, items are retrieved from sequence memory and bound with new positional tags (Figure 4.3, **light blue-grey boxes)**. As in sequence memory, bindings form over time between tags and corresponding items, which are superposed to form representations of specific dependencies (Figure 4.3, **topmost representation, all diagrams**). The timing of each binding

operation is associated with and coordinated by stimulus-related oscillatory activity (see Figure 4.2) (Barczak et al., 2018; Lakatos et al., 2008).

4.7.2 Nonadjacent dependencies

To support efficient encoding, and permit generalization from learned adjacent dependencies, simple nonadjacent relationships may be encoded in an equivalent manner to adjacent dependencies. To accomplish this, all constituent items must be maintained in the sequence buffer long enough to be integrated, and maintained representations need to be selectively propagated to the dependency buffer. This process involves salience filtering (e.g., by repetition suppression, expectancy or attentional processes, reviewed in Summerfield & de Lange, 2014; see Figure 4.3, **centre diagram**), possibly supported by regions such as ventral frontal cortex including the frontal operculum (FO). This is based on findings that the FO appears to be more active during the presentation of infrequent or novel auditory cues (Rong et al., 2018) and responds preferentially to violations of adjacent dependencies (Friederici, Fiebach, et al., 2006). We suggest that this region integrates information from working memory as soon as it is available, but only maintains it over a relatively short time period. In this case, preferential responses to adjacent dependency violations could be explained by FO actively inhibiting representations of both low probability items and short, low probability *n-grams* of contiguous elements (for example, those in which the constituent elements are not suitably ordered).

Salience filtering enables nonadjacent dependencies to be encoded into representations identical to their adjacent counterparts, for any length of dependency fitting into working memory. That is, for any salient **A** and **B** and n irrelevant **X** elements, the sequence **A-X**^{n}**-B** may be rendered into the same encoding as **AB** alone by selectively inhibiting downstream encodings of irrelevant **X** representations (Figure 4.3, **identical top representations, left and middle diagrams**). Discarding irrelevant items results in a more efficient representational code, and in state-transition terms allows for grammaticality judgements to be made on n^{th} order nonadjacent dependencies using only a first-order Markov process.

4.7.3 Hierarchical dependencies

As shown, within our model a dependency comprises multiple superposed *position* $\otimes$ *item* bindings. Although the dependency can incorporate items retrieved from either contiguous or

discontiguous positions in serial working memory, the recoded dependency representation can in both cases be considered a chunk (in Figure 4.3, **left diagram**, the chunk is A-B; in the **right diagram**, the two chunks are $\mathbf{A_1}$-$\mathbf{B_1}$ and $\mathbf{A_2}$-$\mathbf{B_2}$). To encode hierarchical dependencies (Figure 4.3, right diagram, showing nested dependencies), every *dependency* needs to be bound with a unique positional tag, as for individual items. Like individual items, these can also be superposed, forming a superchunk containing a higher-order dependency representation. This process can be recursively repeated to form a single reduced representation of the hierarchical structure of the entire input sequence, integrating progressively increasing amounts of information at higher hierarchical levels of encoding.

The above system is sufficient to compress hierarchical structure into a reduced representation. However, we must encode the reduced representation using more than item positional tags if we are to support unambiguous recovery of specific constituents and comfortably discard the original representations. To avoid generating identical codes for dissimilar structures, we can define sets of positional tags specific to each level of the hierarchy, for example $1_I°$ and $2_I°$ for the first and second items, $1_C°$ and $2_C°$ for the first and second chunks, and so on (see Figure 4.3, **right diagram**). For this reason, it can be computationally beneficial to define all positional tags as convolutional powers (i.e. $base^{exponent}$) of a given *base vector* through repeated self-binding (i.e. $2° = 1° \otimes 1° = (1°)^2$). This, or a similarly invariant function, can be learned by a network such that tags are encoded not just as a function of position (by varying the exponent), but also context (by varying the base vector). Crucially, by binding items only to a finite set of deterministically generated positional tags, the modelled system can always undertake unbinding by re-instantiating the same set of keys; by iterating through every possible tag, all constituent items can be retrieved in sequence.

A given dependency coding network can retrieve and recode items from arbitrary positions in the linear sequence buffer, provided it has sufficient integrative power. Therefore, this representational scheme is capable of encoding not just nested dependencies, but other types of hierarchical dependency such as crossed dependencies. For example, Figure 4.3 (**right diagram**) illustrates retrieval of items 1 and 4 ($\mathbf{A_1}$ and $\mathbf{B_1}$) from sequence memory, which are chunked through binding and superposition; and retrieval of items 2 and 4 ($\mathbf{A_2}$ and $\mathbf{B_2}$), which are likewise chunked. These two chunks form a superchunk representing a nested dependency. Cross-serial dependencies

between words are notably also present in some human languages. Like nested dependencies, these can also be readily represented by our model. We could, for example, have retrieved and bound items 1 and 3, and 2 and 4, respectively, to form a crossed dependency structure. This is illustrated within our coded demonstration (DOI:10.5281/zenodo.3464607). Their inclusion alters the minimum computational requirements for any parsing agent (Stabler, 2004), and thus it is important that a domain-general representational model has the potential to account for them as well as other language-like hierarchical constructions.

Having recoded a linear sequence in terms of its dependencies, it is possible to recover one or more specific items from the representation of hierarchical structure by unbinding using keys that specify context-specific position(s). With reference to Figure 4.3 and using this method: the key $\neg \mathbf{1}_C^\circ$ will recover all position-item bindings of the first chunk; the key $\neg \mathbf{2}_I^\circ$ will recover the second item of every chunk (with each still bound to information on the requisite chunk position); and the key $\neg(\mathbf{2}_I^\circ \otimes \mathbf{1}_C^\circ)$ will recover the second element of the first chunk. This flexible decoding scheme is important because it readily supports manipulation of entire chunks and generalization of dependencies to multiple timescales.

A natural consequence of the above encoding scheme is that the superposition of many dependency representations over time can gradually give rise to a memory trace. This trace will be influenced by the probabilistic distribution of dependencies over the set of input sequences, analogous to implicit learning. It is now possible to computationally model functional characteristics of hippocampal subregions, and by this method it has been proposed that a monosynaptic (entorhinal cortex to CA1) pathway possesses the relevant properties to support implicit learning (Schapiro et al., 2017). This is relevant for sequence learning, where dependencies are established over many trials. Single-trial learning, by contrast, requires processing by further hippocampal subregions. This suggests a prominent role for parts of the hippocampal system in sequence processing, in line with recent findings (Opitz & Friederici, 2004; Schapiro et al., 2017). However, this is not to say that we should expect activation of the hippocampus to be constant throughout. Indeed, the human neurobiology literature suggests that there is a decrease in hippocampal involvement over time when learning an artificial grammar (Opitz & Friederici, 2003), or acquiring a novel semantically meaningful lexicon (Breitenstein et al., 2005). These observations can be interpreted by way of a predictive coding account, in terms of the degree of mismatch between stored and incoming sensory

encodings (Schiffer et al., 2012). We can relate this associative mismatch account of hippocampal involvement to the idea of a superposed memory trace. Namely, any newly presented sequence is likely to be more similar to a superposition of encountered sequences as time progresses and the number of constituents in the superposition grows. Thus, the degree of mismatch, and any activation requisite for such an encoding, is likely to decrease over time.

We propose that hierarchical structure-building is one of the key roles of the dorsal aspect of ventrolateral prefrontal cortex (dorsal VLPFC, incorporating Brodmann Areas 44 and 45). This position is supported by human neurobiological evidence on syntactic processes prominently featuring hierarchical dependencies (Wilson et al., 2017; Friederici, Fiebach, et al., 2006). It has been proposed that at least BA44 supports a recursive, multi-dependency management process, a fact which would explain increases in its activation observed with increasing depths of hierarchical dependency nesting (Friederici, 2011). The recursive reuse of a consistent architecture by VLPFC would be consistent with our recursive use of vector symbolic operations during the encoding of hierarchical dependencies. Repeated superposition of sparse dependency representations will manifest as an increase in local activity as increasing numbers of neurons support the representation; as an example of this effect, consider the increased activity represented at the top of Figure 4.3, relative to the bottom. A "reset" (or re-sparsification) of the buffer will result in a sudden drop in population activity. Such neural accumulation and reset activity has been identified within human intracranial recordings in subjects listening to sentences containing words that accumulate into phrases (Nelson et al., 2017). Furthermore, representations of constituents might not persist beyond the need to encode the reduced dependency representation, a fact that highlights an interesting property of the model: detectable neural delay activity does not need to persist throughout working memory maintenance of the input sequence, consistent with recently reported findings on the neurobiology of working memory (Lundqvist et al., 2018).

It must be emphasised that the flow diagrams in Figure 3 only show *symbolic* information flow. There are likely to be sub-symbolic influences acting over time to hone the associated neural activation patterns, requiring feedback and feed-forward interactions between neural ensembles in the model. These are briefly alluded to in Figure 4.3 as a single exemplar (curved, dashed arrow, **left**) denoting top-down influences acting on sensory cortical representations. These sub-symbolic influences may be investigated further by instantiating a dynamic neural implementation of the

model via the NEF (Eliasmith et al., 2004), where learning algorithms employing, for instance, spike-timing-dependent plasticity, can be applied to learn functions over time (Aubin et al., 2017).

4.8 Unifying neurocomputational accounts: extending the serial order code

Thus far we have described how discrete serial position can be incorporated into neurally and cognitively plausible vector symbolic representations of sequentially organised content by tagging items with their respective ordinality. Within our vector symbolic account of sequence processing, this structure has been suggested as the basis for linear encodings as well as multilevel tree structures within the fronto-temporal language network. However, we have also alluded to the importance of continuous, boundary-relative encodings (*relative position*) in the human brain. In such representations, an item halfway through a sequence would be encoded as having a position of 0.5 relative to the sequence boundaries, for example, a position on a continuous scale that is independent of the absolute length of the sequence.

Rather than leaving the relationship between these encodings implicit, in this section I will set out a concrete description of the computations required to support continuous relative position, which can be understood as extensions or generalisations of the previously described serial order encoding. As will be shown, extending this encoding permits the unification of seemingly anatomically and functionally disparate neurocomputational accounts.

4.8.1 The need for continuous relative position

As described previously, the serial positional encoding of items uses distinct positional tags to mark each item's position within a sequence (i.e. $Sequence = 1^{\circ} \otimes Item1 + 2^{\circ} \otimes Item2 + 3^{\circ} \otimes Item3$). The positional tags are most readily conceived of as absolute ordinal positions (for example 1st, 2nd, 3rd), but this need not necessarily be the case. Indeed, restricting tags to only absolute positions needlessly constrains the explanatory and computational capabilities of the positional encoding.

In neural terms, absolute positional encodings are insufficient to encompass the range of positional representations seen within the brain. Evidence for *relative* serial encodings has been found in primate motor, premotor and dorsolateral prefrontal cortex (Carpenter et al., 2018) and rat hippocampus (Shimbo et al., 2021). Neurons in these areas exhibit activity profiles from which sequential position can be decoded, but in a manner that scales with the length of the sequence. That

is, they exhibit activity that appears to encode position relative to the boundaries of a larger entity, such as the boundaries of a list. How precisely such boundaries are defined remains a matter of interest (Carpenter et al., 2018). However, from a computational perspective, such boundaries can be derived either from bottom-up signals within the stimulus itself, or from top-down expectations about the impending boundary, learned from prior experience (Barczak et al., 2018). Furthermore, such positions need not be defined relative to a one-dimensional sequence in time but could equally be defined in the spatial domain. Such topics are investigated later in this chapter.

As well as supporting the electrophysiological evidence, relative positional encodings demonstrate explanatory added value in behavioural hypothesis generation when compared to absolute positional encodings. For example, in humans, transposition errors are more likely to occur between items at the end of different sized groupings than between the same absolute positions (Henson & Burgess, 1998). Furthermore, using an absolute encoding scheme, the 3rd and 4th items are nominally just as distinguishable from each other in a 5-item list as in a 20-item list. In relative terms, however, these same item positions are much closer together in the 20-item list than in the 5-item list (on a scale of 0 to 1, having positions of 0.6 and 0.8 in the 5-item list, versus 0.15 and 0.2 in the 20-item list). This notional reduction in distinctiveness between neighbouring positional tags for longer lists accords with evidence that humans produce more transposition errors between adjacent elements when recalling longer sequences (see, for example, Haberlandt et al., 2005).

Finally, computationally speaking, relative positional codes have two key advantages. Firstly, they allow dependencies between events or objects to share common encodings despite differences in either the absolute scale or contents of a sequence. This has the potentially enormous advantage of permitting dependency learning to be abstracted across timescales as well as modalities. Secondly, since relative positional codes encode values on a continuous scale (as opposed to discrete ordinal values), it becomes trivial to coordinate them (and thus sequence playback) using any continuous signal, including oscillatory activity.

An account of relative positional encoding is thus highly desirable. Unlike absolute ordinal position, however, encoding relative position using a vector symbolic architecture (VSA) requires us to define a reliable way of transforming continuously varying, real scalar values (technically, any real value between 0 and 1) into distinct positional keys. There are two methods of accomplishing this,

one of which, derived from the VSA literature, is computationally efficient; and the second of which, a novel method, offers explanatory advantages and relaxes neurocomputational assumptions.

4.8.2 Encoding continuous position using fractional binding

For a serial positional encoding to function, each positional tag must be distinct. In VSAs, this is accomplished by ensuring that tags are orthogonal to each other. Since vector symbolic binding produces output orthogonal to both inputs, we can use self-binding to generate new keys in a computation analogous to exponentiation. Thus, as we saw in 4.7.3, it can be useful to define a general formula for positional keys that involves raising a base to an integer exponent, like so:

$$\mathbf{V} = \hat{\mathbf{k}}^{\alpha}$$

This states that the positional key $\mathbf{V}$ computed for the α^{th} value is equal to some unit vector $\hat{\mathbf{k}}$ bound with itself α times. For example, the fourth item of a list would have a positional key $\mathbf{V} = \hat{\mathbf{k}}^{4}$; that is, $\mathbf{V} = \hat{\mathbf{k}} \otimes \hat{\mathbf{k}} \otimes \hat{\mathbf{k}} \otimes \hat{\mathbf{k}}$.

It is difficult to conceive of a "partial" binding operation, such as $\mathbf{V} = \hat{\mathbf{k}}^{0.43}$, in mechanistic terms, and so at first sight this scheme may appear to be restricted to integer exponents (and thus absolute encodings). However, whilst it may seem unintuitive, Plate (1995, 1992) demonstrated that, indeed, the convolutional binding power need not be an integer. Since circular convolution in the spatial domain is equivalent to pairwise multiplication in the frequency domain, it can be demonstrated that the above exponentiation is equivalent to:

$$\mathbf{V} = f^{-1}\left(f\left(\hat{\mathbf{k}}\right)^{\alpha}\right) \qquad \alpha \in \mathbb{R},\ \hat{\mathbf{k}} \in \mathbb{R}^{n}$$

where f and f^{-1} are the forward and inverse Fast Fourier Transform (FFT) respectively; α is any real number; and as before $\hat{\mathbf{k}}$ is a unit vector, and $\mathbf{V}$ the resulting positional key vector.

In this way, the scheme we defined previously for serial encoding can be used to encode items by some continuous measure of position, including relative position, bringing it in line with the neural and cognitive evidence. Furthermore, and remarkably, by binding multiple continuous position encodings together, continuous trajectories in 2D, 3D or higher-dimensional space can also be represented (Plate, 1992), allowing such a system to map visuospatial domains, for example. However, this particular method formalises some constraining assumptions; we can, in particular,

relax the requirements for either Fourier domain computations or convolutional powers by defining a more general transformation. This novel vector symbolic mechanism is outlined below.

4.8.3 Encoding continuous position using basis functions

The above method uses fractional convolutional binding to encode orthogonal positional keys along a continuous trajectory. However, there is no neurocomputational need to restrict orthogonal codes to those produced by exponentiation. Neither is there a formal requirement that resultant codes be uniformly spaced along a trajectory, as they are using the fractional binding method. One can abstract away from such concerns using the following novel vector symbolic formula:

$$\mathbf{V}_{\alpha} = \sum_{i=1}^{n} \psi_i(\alpha)\, \hat{\mathbf{k}}_i$$

where α is, as above, any real number; n is an integer number of orthogonal encodings; ψ is a set of n basis functions spanning the real-valued input domain; $\hat{\mathbf{k}}$ is a set of n distinct unit vectors, and $\mathbf{V}_{\alpha}$ is again the resulting positional key vector. Summation here denotes superposition of its operands.

The formula is sufficiently compact that a prose explanation is assistive. Here, each distinct basis vector $\hat{\mathbf{k}}_i$ is scaled by the output of a corresponding basis function ψ_i. The basis vectors $\hat{\mathbf{k}}$ must be orthogonal for effective encoding, and could be a set of integer binding powers as described previously. However, they need not be; since any two randomly selected high-dimensional vectors are likely to be approximately orthogonal, this could be a set of random vectors. Scaling occurs by multiplying every component of the vector with the scalar output of the given basis function. These scaled vectors are subsequently superposed, producing one key vector $\mathbf{V}$ that varies continuously with the real-valued input scalar, α. Thus, a real valued scalar α is transformed into a continuously changing vector $\mathbf{V}_{\alpha}$ by summing a mixture of vectors, using α to control the mixing proportions. The resultant key vector $\mathbf{V}_{\alpha}$ can serve as the input for further vector symbolic operations, as with the fractional binding or absolute positional encoding methods described previously.

Casting off the nomenclature, the method is readily portrayed in visual terms, as in Figure 4.4, which illustrates the basis function method applied to a scalar α derived from the phase of some upstream input. Basis functions such as evenly spaced Gaussian curves, as depicted, can mimic the fractional binding system described previously, but since these could in fact be any set of functions

spanning the input domain, including heterogeneous functions that non-uniformly span the input domain, the basis function method is less assumptive than the fractional binding method.

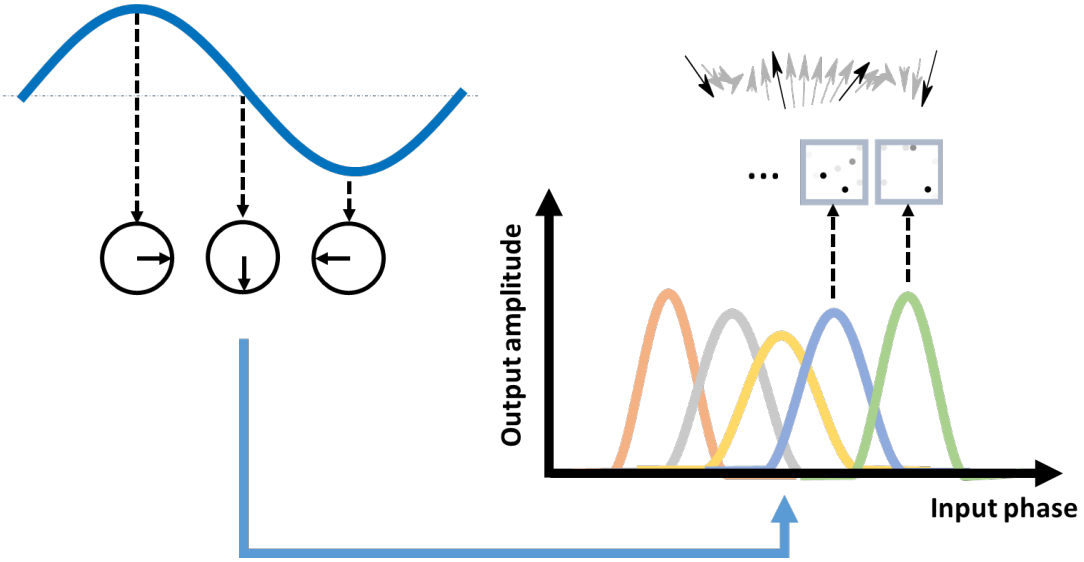


Figure 4.4 **Transforming continuous phase into discrete positional encodings**. Within a position-encoding population, some vector components (and thus neurons) respond preferentially to specific values of input phase. Thus, at any given phase, a subset of neurons will activate. This is equivalent to activating one of a finite number of sparse, distributed positional encodings (boxed dots, representing basis vectors) based on tuning to specific values of phase (coloured Gaussian curves, representing basis functions). Randomly selected sparse, distributed vectors are approximately orthogonal (black arrows). Moving through continuous values of phase causes the output to morph smoothly between the orthogonal representations (grey arrows), forming a continuous trajectory.

A key demonstration of the utility of non-uniform basis functions can be seen in a novel simulation of behavioural performance in a free recall task. To store an n-element sequence $\mathbf{S}$, as previously described, each element $\hat{\mathbf{e}}_j$ is bound with its respective relative positional key $\mathbf{V}$ and superposed in a buffer, like so:

$$\mathbf{S} = \sum_{j=1}^{n} \hat{\mathbf{e}}_j \otimes \mathbf{V}_{(j-1)/(n-1)}$$

where the keys are derived from the basis function method above and the subscript $(j-1)/(n-1)$ simply denotes the relative position of the j^{th} element in n (i.e. the first element has a relative position of 0, and the last element a position of 1).

Likewise, as previously described, recall of a specific item $\hat{\mathbf{e}}_j$ can occur by unbinding using the appropriate positional key, thus:

$$\hat{\mathbf{e}}_j = \mathbf{S} \otimes \neg\mathbf{V}_{(j-1)/(n-1)}$$

where $\mathbf{V}$ can be derived, as before, solely from the appropriate basis functions and relative position. This provides a foundation for a vector symbolic model of free recall using relative positional encodings. Here, the basis function method provides notable added value by way of the following configuration.

If, rather than spacing the maxima of basis functions uniformly, as in Figure 4.4, they are spaced non-uniformly, along a sigmoid curve, one presupposes that positions at the "edges" of things are better encoded than those in the middle. Likewise, the maxima of the basis functions can be non-uniform such that keys at the edges are always weighted more strongly than those at the centre. These two non-uniformities can be seen in the idealised basis functions shown in Figure 4.5. This can be combined with a simple decay constant that scales older items towards zero as new items are added, producing a relatively parsimonious vector symbolic simulation of primacy and recency in terms of the required dynamic interactions (see Figure 4.6). In this simulation, we used a modified variant of the ideal basis curve scheme shown in Figure 4.5, in which tuning was far more sparse and not entirely symmetrical, with clearer encoding of the most recent items compared to the earliest, and which was subtly randomised between participants (see Figure 4.6 legend). The result is a convincing similarity between the real and simulated data (c.f. Figure 4.6 left and right panels; see Murdock, 1962). However, further fine-tuning parameters in this model would improve its concordance with other neuroimaging and behavioural datasets, although this is beyond the scope of this chapter. However, it should be evident that the basis function method of continuous position encoding is extremely versatile. This is before one considers its strengths in integrating disparate neurocognitive perspectives, the focus of the next subsection.

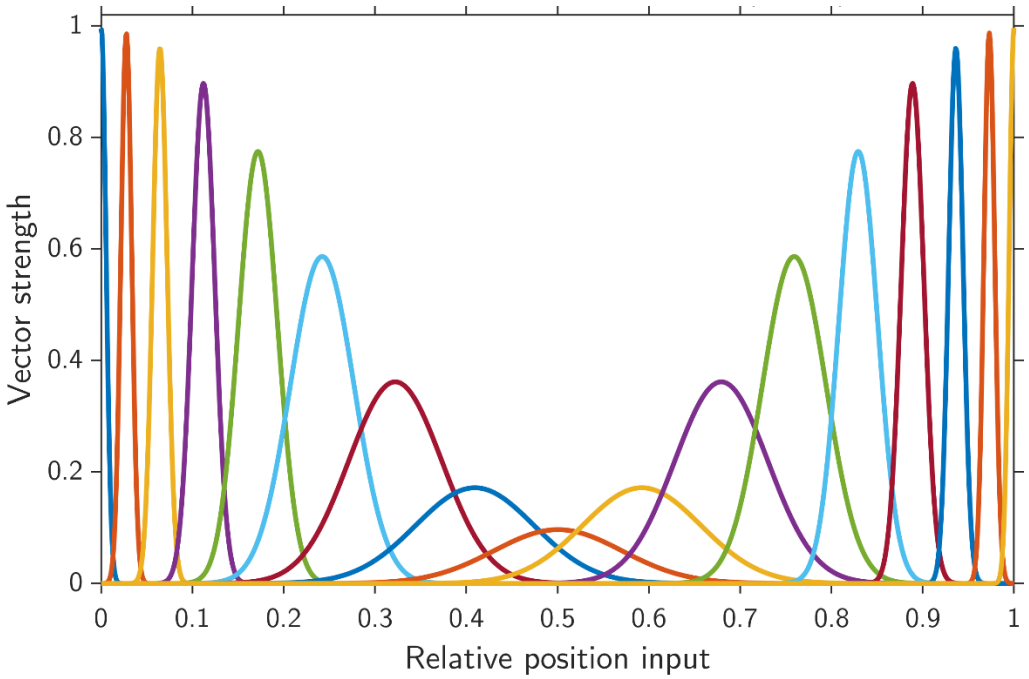


Figure 4.5 **Non-uniform basis functions for relative position encoding.** These particular basis functions generate boundary encodings that are both stronger and more precise than those at the centre. The figure is sufficient to characterise the $N = 17$ basis functions, but for completeness, these are scaled Gaussian curves ($\sigma_{edge} = 0.005$, $\sigma_{centre} = 0.07$; $scale_{edge} = 1$, $scale_{centre} = 0.1$) where the N mean, standard deviation and scaling factors were themselves determined by reading from N equally spaced points on a Gaussian distribution: $\mu = 0.5$, $\sigma = 0.15$ for the standard deviations and scale factors; and $\mu = 0.5$, $\sigma = 0.3$, cumulatively summed to produce a sigmoid curve for the means. Here there are 17 unique positional keys, but more or fewer can be flexibly generated by following exactly the same scheme. This flexibility can be useful when simulating the impact of variable key count within a virtual population of subjects, for example (that is, we do not need to assume that positional encoding fidelity is identical in every virtual subject of the population, so each subject can use a different number of basis vectors).

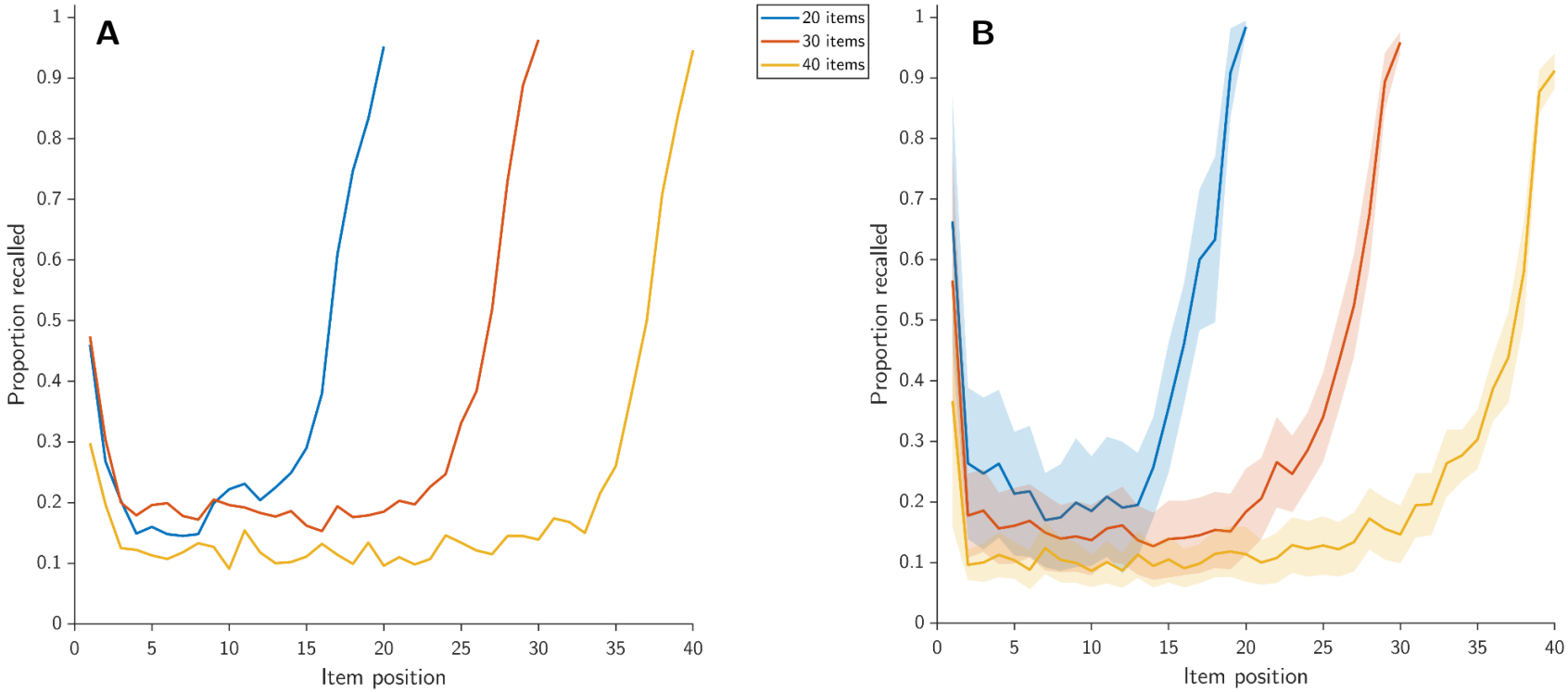


Figure 4.6 **Simulation of primacy and recency effects in free recall, 20 to 40 items**. As in Murdock (1962), N = 20 distinct participants were tested per list length. As in the original experiment, each participant was exposed to 80 lists and recall proportion computed from these. Shown are group mean proportions of items recalled in specific positions under the original task (**A**, *left*) and a simulation of free recall (**B**, *right*; shading depicts CI_{95}). Here the *basis function* method of relative position encoding was used to encode lists. Items were stored in a single (180-dimensional) vector symbolic buffer, unlike existing methods showing similar behaviour (for example, Choo & Eliasmith, 2010, which requires two buffers). Every time an item was added to the buffer by superposition, or recalled from it by unbinding, the existing buffer was multiplied by a decay constant (here 0.94). Recall yielded a correct/incorrect item as in real life, from which performance was calculated. Basis functions used to encode position were non-uniformly spaced and scaled, so that positional codes close to the middle of the list were less distinct than those at either end. The number of distinct positional keys available to each virtual subject was variable, uniformly sampled at random in the range (15, 25). Centres of tuning curves were positioned along a Gaussian distribution with random sigma in the range (0.5,0.55). Curves centred at exactly 0 (beginning of list) were disallowed, but not those at 1 (end of list), producing a bias towards high-fidelity encoding of the most recent items, but not the oldest.

4.8.4 Relative position encodings in accounts of perception and memory

As has been shown, a vector symbolic instantiation of discrete serial position can be readily extended to support continuous relative positional encoding, reflecting evidence in the literature and relaxing neurocomputational assumptions. Namely, the encoding scheme relaxes assumptions of absolute ordinality, the uniformity of basis vector mappings, and the requirement for convolutional powers. This relative positional encoding can also support evidentiary triangulation across cognitive domains, because it can be incorporated into a wider range of neurobiologically plausible mechanisms. This subsection provides evidence for this by exploring the role of the vector symbolic relative positional encoding within novel models of stimulus-driven speech segmentation and hippocampal replay. Because these models utilise a common encoding scheme, they are amenable to integration into a broader view of sequence processing.

Let us first address speech segmentation. A neurobiologically plausible, vector symbolic account of stimulus-driven speech segmentation is useful for two reasons. Firstly, such an account is required in order to explain, from first principles, how the vector symbolic sequence encodings described above can be derived from stimulus encodings known to arise in sensory cortex. Secondly, by codifying models of higher-order dependency encoding (see section 4.7) and stimulus-driven speech segmentation within a single conceptual framework, a VSA model brings us closer to reconciling an account of "bottom up", stimulus-driven segmentation with evidence of the crucial role played by ventrolateral prefrontal cortex in human auditory segmentation (discussed previously in 1.4.1, see page 19).

Here, for the first time, a vector symbolic account of speech segmentation is proposed. Key assumptions of the model are derived from an existing, primarily descriptive, account of speech segmentation in primary auditory cortex (A1) provided by Giraud & Poeppel (2012). Within the Giraud and Poeppel model, "bottom-up", stimulus-driven signals arising in A1 input layer IV coordinate low-frequency oscillatory activity (for example, theta band oscillation, assumed to arise in superficial layers of auditory cortex), which in turn modulates high-frequency (gamma band) activity, guiding the segmentation of the input stream into discrete packets of temporally organised spike-trains (arising in output layers II/III of auditory cortex) that are fed forward from A1 to higher order areas. The Giraud and Poeppel model is based on a synthesis of evidence across studies. For example, stereotactic EEG studies in humans have demonstrated that the power spectrum of

responses to speech in A1 strongly correlates with the power spectrum of the stimulus envelope in two critical frequency bands, theta and gamma, which are coupled together in a nested relationship during speech presentation (Liégeois-Chauvel et al., 2004; Giraud & Poeppel, 2012). Likewise, theta band tracking of the speech envelope has been shown to be critical to speech intelligibility (Nourski et al., 2009; Luo & Poeppel, 2007; Ahissar et al., 2001).

The VSA model of speech segmentation is illustrated in Figure 4.7. Here, stimulus-driven envelopes, derived from responses arising in the auditory periphery, can preferentially drive frequency-specific envelope-tracking oscillators (Figure 4.7**A-B**). A low-frequency oscillator is a source of relatively stable phase information (Figure 4.7**C**). This low-frequency wave contains information sufficient to identify boundaries in a continuous signal, resetting each time a perceptual boundary is reached. However, to increase the fidelity of the model, this stimulus-driven activity could also be modulated to varying degrees by learned top-down influences (Barczak et al., 2018).

Instantaneous phase can be extracted by a variety of methods but here, during spiking neural simulation of this model, was extracted by approximating the Hilbert transform using neurobiologically plausible operations (see Appendix 2: Supplementary figures, Figure 0.10). Instantaneous phase can be used as a repeatedly resetting, but otherwise monotonically increasing source of positional information using the relative positional encoding scheme described in subsection 4.8.3, creating a continuous trajectory of positional codes that describe the current position of, for example, a phoneme within the wider context of its containing syllable (Figure 4.7**D**). This relative position can be bound with a representation of the recent auditory history (for example, a snapshot of neural cepstral coefficients; see Bekolay, 2016) to encode an ordered list of content within context (for example, a list of phonemes within a given syllable, or of syllables within words; Figure 4.7**E-F**). In summary, a continuous stream of auditory input is transformed into a temporally organised vector symbolic encoding of nested, discretised phonetic information.

Studies in the human EEG and ECoG literature continue to provide insights that accord with this vector symbolic model. For example, it appears increasingly likely that linear, temporally ordered phonetic encodings, the immediate output of this model, initially arise in superior temporal gyrus (STG; Brodmann Area 22), in areas overlapping with Wernicke's area (Wernicke, 1874, 1881), lesions to which classically result in Wernicke's aphasia (or *receptive aphasia*), chiefly characterised

by marked difficulties comprehending language. Electrophysiological recordings in middle-to-anterior STG and posterior STG have revealed stimulus-driven activity that suggest integration of temporal and phonetic content across syllables and phrases, respectively (Yi et al., 2019; Giraud & Poeppel, 2012). Furthermore, recent ECoG evidence suggests that both spectral and temporal cues in speech are indeed encoded in STG using a common spatial neural coding scheme (Fox et al., 2020), akin to the population coding of bound items posited by the vector symbolic segmentation model.

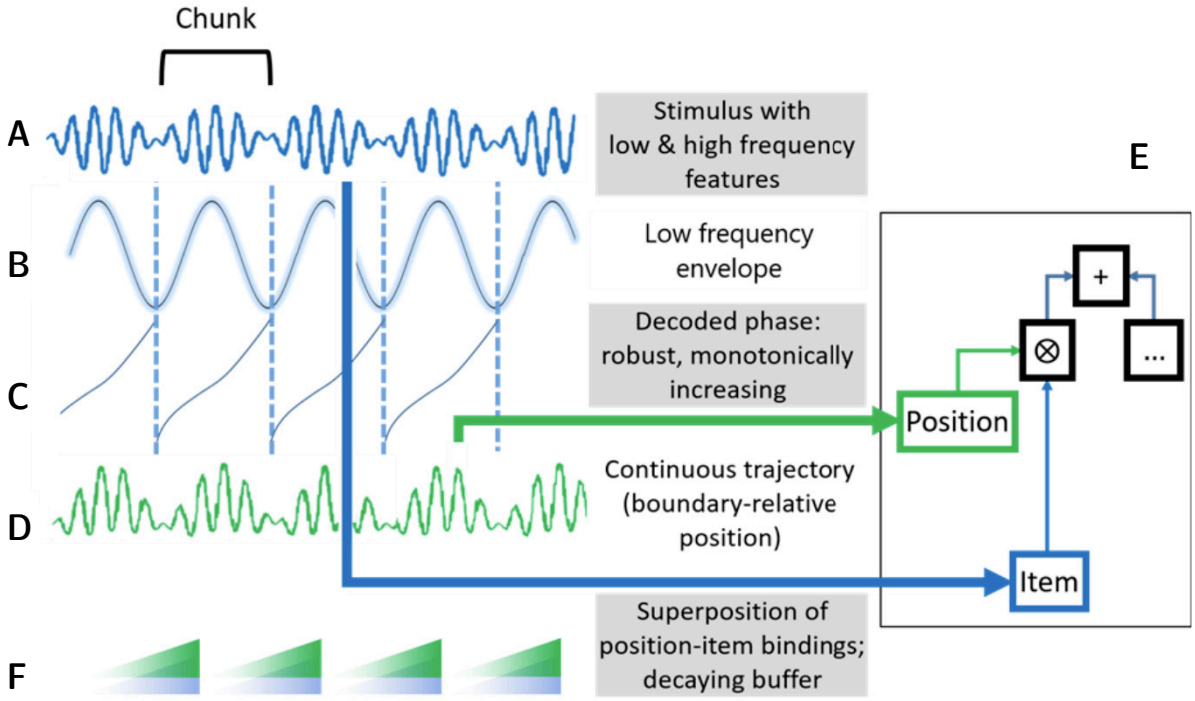


Figure 4.7 **A putative mechanism for deriving ordinal serial encodings from "bottom-up" segmentation of auditory input**. A low frequency envelope tracking the stimulus acts as a stable source of phase information (**A**-**B**). Because phase produces a monotonically increasing signal within each oscillatory period, but resets at the boundaries of the period, it acts as a source of unique chunk-relative positional information (**C**). A continuous trajectory of positional encodings can be driven by this phase signal alone (**D**). By conjunctively combining the instantaneous positional encoding with the instantaneous stimulus encoding (**E**), and periodically superposing these conjunctive representations, an ordinal serial encoding of discrete perceptual items can be generated from the original continuous stimulus encoding. This encoding likely decays over time (**F**).

Thus far, the proposed models codify a broadly feedforward, neurobiologically plausible route of information flow, illustrating how higher order dependency structure may be derived from continuous sensory input. Specifically, the described models show how representations may develop as they pass from the auditory periphery to primary auditory cortex, auditory association cortex, and onward to prefrontal cortex, notably ventrolateral prefrontal cortex. However, in this account, there has been scant mention of the feedback influences that must necessarily factor into an account of sequence processing, in particular of those networks supporting mechanisms for sequence prediction, generalisation and memory. These concepts would appear to fall squarely within the remit of the hippocampus (Buzsáki & Tingley, 2018; Davachi & DuBrow, 2015), and it is therefore notable that online involvement of the hippocampus appears integral to sequence processing (as previously discussed in section 4.3).

There are myriad existing computational models describing the hippocampus as a whole (see, for example, Lisman & Jensen, 2013; Schapiro et al., 2017; Whittington et al., 2020), but vanishingly few VSA models (actually, to date, only one: Trujillo & Eliasmith, 2014, although there have been attempts since to use VSA techniques to specifically instantiate grid cell encodings; see Dumont, 2020). Furthermore, there are no VSA hippocampal models that have been designed from the outset to coordinate with other brain networks in a neurobiologically and cognitively plausible manner. With this in mind, a novel vector symbolic model of the hippocampus is explored below in detail.

We can model the hippocampus by conceiving of it as a sequence playback engine in which playback is centrally coordinated by oscillatory activity. Although the focus here is primarily auditory sequences, these might just as well be sequences of spatial locations, motor events or emotions, and thus the mechanisms here may be considered domain-general. In this case, the model's agnosticism with respect to the form of the input concords well with suggestions that the hippocampus treats both time and space in a similar manner (Buzsáki & Tingley, 2018).

The model is illustrated and described in Figure 4.8. A sequence (for example "ABCD") is firstly stored during memorisation as a vector symbolic list encoded using relative position (as described in subsection 4.8.3). Subsequently, this sequence can be traversed by using a continuous position encoding driven by an oscillator. To be precise, the stored sequence is unbound using a positional key, where the positional key is derived from the phase of an oscillator using the basis

function method. The coordinating role of hippocampal theta oscillations is by now well documented, both within the hippocampus and between the hippocampus and external regions such as prefrontal cortex (Buzsáki, 2002; Lisman & Jensen, 2013; Solomon et al., 2019; Zielinski et al., 2019). In this model, phase derived from the theta oscillator produces a sawtooth wave that, when used to generate a positional key, produces repeated past-to-future scanning through the stored sequence. The generated signals also exhibit phase-amplitude coupling, observed as theta-gamma coupling in empirical hippocampal data (Lisman & Jensen, 2013). By adding to the sawtooth-shaped phase wave an offset representing the subject's egocentric position in space or time (Figure 4.8, in red), one can simulate scanning that passes from the present position into the future. That is, one can simulate the prediction of upcoming sequence items given the present spatial location (for example, points in a maze, if storing a series of spatial locations) or temporal location (for example, time offset, if recalling a list of words over time). The scanning mechanism generates cycles of positional vectors that correspond well with accounts of place cells (Moser et al., 2015). Meanwhile, the steady advancements in activation observed as the subject passes through different locations correspond to empirically observed phase precession (O'Keefe, 1976).

Note that the features of this novel hippocampal model by no means presently exceed those addressed by all state-of-the-art alternatives. For example, Dumont & Eliasmith (2020) have simulated hippocampal function in spiking neural terms, generating rich two-dimensional place- and grid-cell maps that exceed the complexity of place cell activity presently simulated under our model. Similarly, Whittington et al. (2020), in their rich model, the Tolman-Eichenbaum Machine (TEM), have not only provided a coherent account of hippocampal function at the neural level, showing that their model learns realistic grid- and place-cell mappings, but have also used these principles to simulate the learning of both spatial relations and relations between abstract concepts. Specifically, the TEM implements an algorithm that supports *transitive inference*, whereby exposure to relationships between specific items permits inference of new, unseen relationships, supporting sensory predictions. This is an important feature, since it defines a role for the hippocampus in abstracting across chained linear and hierarchical relations, allowing it to make predictions on the basis of prior knowledge even in novel circumstances. We do not presently address learning or transitive inference in this hippocampal model. However, our model does have key features of its

own, including: relative simplicity; the fact that it explicitly specifies a role for oscillatory activity; and that it utilises the same coding principles as other diverse phenomena addressed in this chapter.

In summary, in this section, vector symbolic relative position encodings have been explored as a way of triangulating and unifying seemingly disparate empirical findings. The sum total of all the hitherto presented region-specific models, VS-BIND, at this point proposes testable mechanisms concordant with evidence from regions across temporal and prefrontal cortex, including sensory and association cortex, medial temporal lobe, dorsolateral and ventrolateral prefrontal cortex.

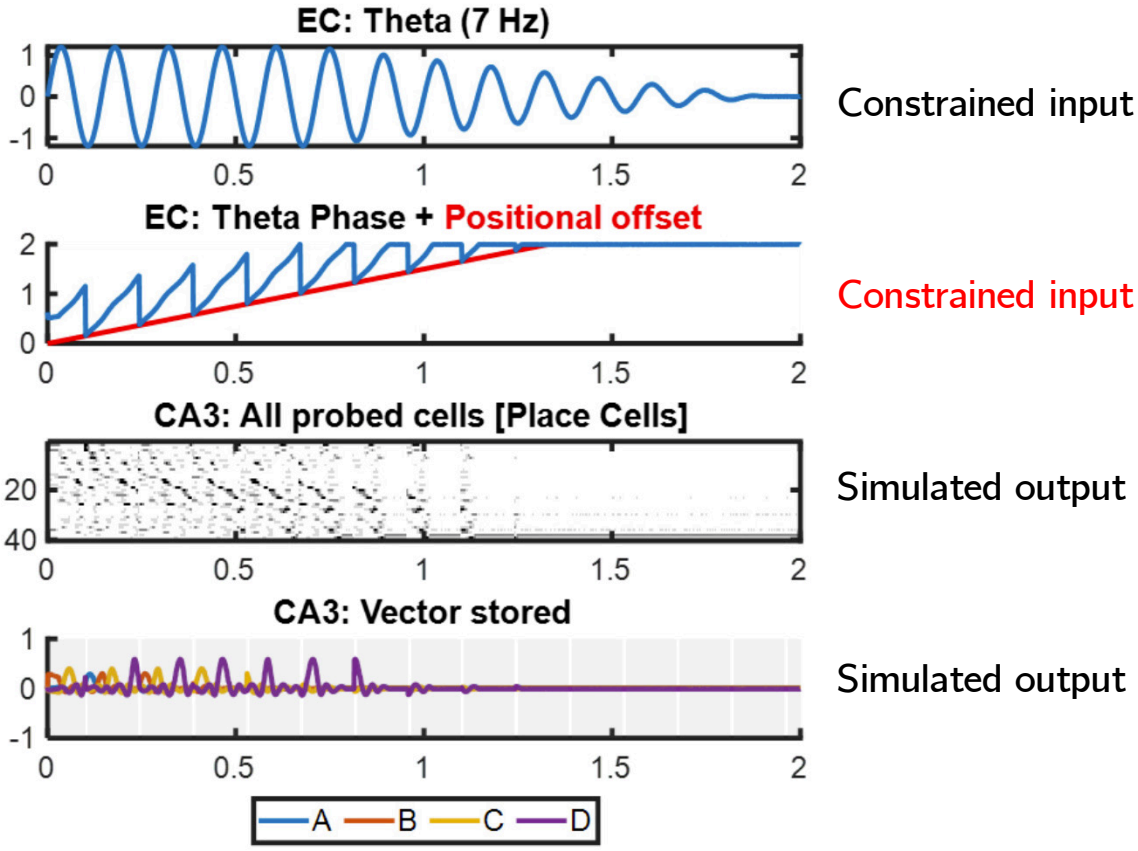


Figure 4.8 **Utilising continuous positional encodings within a model of the hippocampus**. The above shows simulated hippocampal activity, depicting the repeated replay of a predefined vector-symbolic sequence "ABCD". Here, scalar theta oscillations (top row) are an input to the model that act as a coordinating influence on the replay/pre-play process. Though the oscillating signal is provided as a manual input, it could have been generated trivially by a feedback loop between two simulated neural populations. The theta signal drives an encoding of phase (second row, in blue) that can be manipulated by simple addition or subtraction of a positional input to the model (second row, in red) to suggest movement towards or away from a goal state, respectively. This state may be spatial, as in the end of a maze, or more conceptual and temporal, such as "the end of an auditory sequence". When oscillations are strong, the replay/preplay process is active. When theta diminishes, playback ceases. Here, theta and the scalarpositional offset drive activation of a continuous positional vector encoding (third row, depicting a raster plot of simulated output from the model, visualising a random subset of the dimensions of the population code representing position). Because of the sawtooth nature of phase, the generated positional population vector scans repeatedly through some of the same positional code. Finally, the previously stored sequential representation "ABCD" is scanned using the generated positional code to recall individual items (bottom row, showing the correlation of the simulated vector output with each possible item vector "A", "B", "C" and "D", at every time point). Replay and preplay of upcoming sequence items can be observed in the simulated output (bottom row), recapitulating phase precession. We might posit that the stored representation being scanned (i.e. "ABCD") should be activated by environmental context, which would allow for an unfathomable variety of recall sequences.

4.9 In conclusion: Predictions emerging from the structure of VS-BIND

Vector symbolic operations implemented in Artificial Neural Networks, as we have established, have the potential to further our understanding of combinatorial binding at neural, cognitive and behavioural levels, generating site-specific neural correlates ripe for testing. Testing falsifiable predictions made by this model is important to provide evidence about the plausibility of the combinatorial operators and processes modelled here.

Our model, VS-BIND (Vector-symbolic Sequencing of Binding INstantiating Dependencies) does not exist in a vacuum and occupies a landscape of alternative models, a number of which we have introduced here. However, one of VS-BIND's key strengths is that it draws together existing computational explanations under a common framework and in a shared, consistent language. To illustrate this point, Table 4.1 enumerates key features of a number of alternative models or datasets discussed in this chapter, and lists those addressed by VS-BIND here. Whilst it can be seen that some individual models have explanatory capabilities that are presently beyond those of VS-BIND, it can be appreciated that our model applies consistent principles to explain a variety of phenomena that have not been reconciled in existing neuro-computational models.

The principles followed under our approach produce a number of emergent predictions with respect to specific phenomena, enumerated in Table 4.1. More generally, however, the components of VS-BIND we have outlined here suggest the following predictions:

- **Superposition of sparse vectors** manifests as steadily increasing net activation (Nelson et al., 2017).
- **Binding produces orthogonal vectors** by re-coding, which might be observed as reductions in neural responses, for example reductions in high gamma activity during neural local field potential recordings (Nelson et al., 2017).
- **Chunking operations** are intrinsic to linguistic bracketing and both produce and depend upon dynamic patterns of both increasing and decreasing oscillatory activity (Ding et al., 2016).

- **Relative ordinal position** is a crucial neural code when binding serial input in artificial and biological neural systems, indicating that time-sensitive or serial-position-sensitive cells, wherever they reside (Carpenter et al., 2018) are indispensable for complex combinatorial binding in sequence processing.
- **Hippocampal system involvement** is not only required for recurrent activity to establish associations during Hebbian learning, but components of this system are also involved in structured sequence processing, establishing rules and dependencies across temporal scales through interactions with at least inferior frontal cortex.

In summary, the VS-BIND model describes how sequences can be processed and represented using established VSA representations and operations, with the main goal of generating falsifiable predictions of neural correlates. While it was beyond the scope of this account to model the learning processes, VSA-based models such as ours could now be extended with spike-timing-dependent plasticity or other available processes.

It is also important to note that, although our model can represent novel sequences of arbitrary items, the model does not bind arbitrary items together, which risks producing uninterrogable or unrecoverable bound representations. Instead, VS-BIND only uses the binding operator to combine an item with a known positional or contextual tag. Critically, these positional tags are known to the model, being drawn from a small, stable set of codes, deterministically re-instantiated by dlPFC neurons or hippocampal "time cells". This process allows retrieval of arbitrary items without the need for modeller-specific knowledge or a look up table: all that is needed is the set of positional tags (re-instantiated by dlPFC or hippocampus) and the sequence representation. Such models could, in future, be combined with alternative systems that explicitly support the binding together of arbitrary items, which for example support the flexible manipulation of linguistic relations in working memory (Hummel et al., 2004).

Table 4.1: **VS-BIND in the landscape of alternative models.** This table outlines key phenomena that can be encapsulated by the combinatorial and positional coding mechanisms outlined in this chapter. VS-BIND is compared to notable alternative models (M) and one empirical dataset (E).

Modelled (*M*) or empirical (*E*) constraints	Lisman & Jensen, 2013 (*M*)	Giraud & Poeppel, 2012 (*M*)	Murdock, 1962 (*E*)	Doumas *et al.*, 2008 (*M*)	Whittington *et al.*, 2020 (*M*)	VS-BIND	**Emergent predictions of VS-BIND**
Site-specific mechanistic constraints on neurocomputations	•	•			•	•	- dlPFC, vlPFC and the hippocampus integrate relative ordinal position within relational codes
Hippocampal	•				•	•	- Phase precession results from addition of a perceived egocentric positional offset vector to a theta-phase-derived relative positional code
Theta-gamma phase-amplitude coupling	•					•	
Generalisable mechanism (encode position in space/time/abstract domain)	•				•	•	
Seeks to explain place cell activity	•				•	•	
Seeks to explain phase precession	•					•	
Auditory segmentation		•				•	- Discretely encoded auditory items are bound with ordinal positions defined relative to the minima of the low-frequency envelope
Low-frequency phase drives segmentation		•				•	
Cortical layer-specific predictions		•					
Free recall			•			•	- The primacy and recency effects result in part from the tuning curves of specialised boundary-selective neurons coding for relative position
Primacy and recency effects			•			•	
Recency effect stronger than primacy effect			•			•	
Confusion of items increases with list length			•			•	
Relational coding				•	•	•	- Linear and hierarchical dependencies are represented in vlPFC using one or more relative ordinal positional codes to highlight salient items
Hierarchical encodings					•	•	
Transitive inference					•		
Relational learning from Hebbian principles				•	•		
Spanning domains above						•	- Boundary-relative positional codes, forming a continuous trajectory through orthogonal population states, underlie operations in a host of cognitive domains

To conclude, we have presented a blueprint for a neurobiologically plausible computational model, VS-BIND, outlining its principal mechanisms for encoding sequence dependencies. We have also highlighted aspects of VS-BIND that support ongoing efforts to simulate cognitive functions relevant to segmentation, chunking, recall and prediction using VSAs within spiking neural networks. In summary, as outlined in our initial aims (see section 4.2), we have sought to motivate a coherent computational perspective on sequence processing that unifies a diversity of hypotheses traditionally beyond the remit of any single model, and to use this perspective to constrain emergent predictions on brain function. We have accomplished this by reviewing a diverse literature and proposing novel, computationally specific vector-symbolic mechanisms by which key cognitive operations can be achieved, including a novel relative positional encoding method (see 4.8.3). In summarising key tenets of VS-BIND, Table 4.1 provides an illustration of the extent to which we have achieved our aims, setting out our model in the landscape of alternatives and summarising key emergent predictions. Further developments of VS-BIND, in conjunction with other models (such as those that use classically compositional approaches aimed at modelling language-specific or linguistic properties), carry tremendous potential to better understand fundamental aspects of cognition and to guide the pursuit of neurobiological correlates of complex mental structures.

4.10 Data accessibility

Supporting MATLAB code can be found at doi:10.5281/zenodo.3464607 (Calmus, 2019).

The work presented in this chapter was primarily conducted by the author, except where explicitly stated. The author designed and implemented the sequence learning task in MATLAB following discussion with academic supervisors. The medical and surgical team at Iowa University Hospitals and Clinics undertook clinically indicated assessment, imaging and implantation of intracranial electrodes in human neurosurgical patients. Members of Iowa University's Human Brain Research Laboratory (HBRL) confirmed contact coordinates and regions of interest through co-registration and parcellation of structural imaging. The task was executed by our on-site colleague Zsuzsanna Kocsis, to whom we remain indebted, or by the author and HBRL staff members during collaborative visits. Subsequent electrophysiological analysis and writing were conducted by the author. Academic supervisors provided additional advice, comments and discussion.

5.1 Abstract

Sequences unfolding over time contain relationships between neighbouring and distant items that are often critical for information extraction. The ability to perceive this structure is relevant to language, music and other non-linguistic cognitive domains, and has been associated with a network of key fronto-temporal regions, including inferior frontal gyrus (IFG) and frontal operculum. These regions seem to detect statistical regularities and, ultimately, use them to organise sequential input into mental structures. However, there is a pressing need to clarify the mechanisms by which they achieve this, and the neural codes involved. Artificial Grammar Learning (AGL) tasks can provide clarity on this process. In an AGL task, subjects are exposed to sequences conforming to a grammar, and then tested on their ability to distinguish novel "grammatical" and "ungrammatical" sequences.

We previously analysed electrocorticographic (ECoG) data collected in 12 human neurosurgical patients undertaking an auditory AGL task (the "AxB task", Chapter 3), using univariate analyses to reveal engagement of the fronto-temporal language network. We also previously outlined a computational model of sequence processing, VS-BIND, triangulating diverse findings to generate consistent and falsifiable neurocomputational hypotheses. In order to test assumptions of this expansive model using real data, however, we ideally require methods of analysis capable of revealing the structure and flow of isolated representations. Multivariate analyses aggregate neural information across multiple sites and times into high-dimensional spatiotemporal vectors containing more information than can be seen using mass univariate testing. In this study, I undertook a battery of multivariate analyses, including completely novel analyses, to reveal effective functional connectivity between regions of interest; critical aspects of regional syllabic encoding; time-resolved representational signals; and causal flow of relatively specific forms of information across the brain.

This study replicates and advances previous findings that subregions of IFG and frontal operculum have markedly different roles in sequence processing, and suggests that representations in IFG pars triangularis and precentral gyrus may integrate ordinal information, as described in VS-BIND. It also suggests that, whilst auditory sequences elicit feed-forward signals emanating from primary auditory cortex to other areas, expectation-driven predictions feed back to regions including primary auditory cortex, causing a cascade of prediction-error interactions. These findings are consistent with the predictive coding hypothesis, which suggests 'high-level' predictions propagate to 'low-level' regions.

5.2 Introduction

The human brain can learn relationships between complex stimuli distributed over time, but we still have a limited understanding of the mechanisms by which it accomplishes this, the nature of the learned representations, and their neurobiological substrates. A salient example of this ability is our facility with language, where a knowledge of syntax allows us to reconcile structurally associated words and phrases that may be clustered together or separated by intervening content (Tettamanti & Perani, 2012). This knowledge is obtained implicitly during development and in spite of the fact that linguistic tokens must be extracted from a continuous speech stream where acoustic boundaries are insufficient cues to input segmentation (Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996; Aslin et al., 1998). To accomplish this, it appears that humans (and, to varying extents, other species; Sainburg et al., 2019; Wilson et al., 2018; Wang et al., 2015; Wilson, Kikuchi, et al., 2015) have developed an innate sensitivity to the statistical co-occurrence and ordering of stimulus elements. This has been demonstrated using a class of statistical learning task known as artificial grammar learning (AGL) tasks, in which participants are first presented with previously unseen training sequences that all implicitly follow a set of undisclosed rules (a grammar), and subsequently tested on their ability to distinguish novel grammatical from novel ungrammatical sequences. Learning has been established across stages of development (infancy, childhood, adulthood; see for example Reber, 1967; Saffran, Aslin, et al., 1996; Saffran et al., 1999), following mere exposure after a matter of minutes, using stimuli in a variety of domains (Henin et al., 2021; Milne et al., 2018; Stobbe et al., 2012; Saffran et al., 1999), and often implicitly, without awareness of what has been learned. We ourselves demonstrated reaction time effects under an AGL task run on healthy adults (see Chapter 2), finding a significant difference between reaction times to grammatical and ungrammatical sequences across all participants, even in the subset of participants who were unable to explicitly identify the relationships.

What exactly is learned under these tasks still requires clarification, but there is strong evidence that language and non-linguistic statistical learning tasks engage not only domain-selective networks including sensory cortices, but also a domain-general, primarily fronto-temporal network of regions, including ventrolateral prefrontal cortex (inferior frontal gyrus and frontal operculum; Hertrich et al., 2020; Friederici, 2020, 2011), and the hippocampus (Schapiro et al., 2014; Covington et al., 2018), that collectively represent sequential stimuli at different levels of temporal abstraction (Ding

et al., 2016; Buiatti et al., 2009) and differentially engage depending on the complexity of the sequencing relationships (Wilson et al., 2017; Friederici, Fiebach, et al., 2006). Additionally, it has long been known that areas including the posterior temporal lobe and temporoparietal junction (TPJ; including angular gyrus and supramarginal gyrus) are critical for language (Wernicke, 1874, 1881). Whilst ventrolateral prefrontal cortex is thought to initiate responses to syntactic or proto-syntactic sequencing violations, the TPJ is known to be recruited in response to both semantic and syntactic violations (Friederici, 2011, 2012). It has therefore been suggested that this area is involved in semantic-syntactic integration.

Despite the need for clarity regarding what is learned under AGL paradigms, it is known that these tasks elicit encoding of a plurality of relevant features including identities of syllables and learned word-like units, ordinal representations, and transition probabilities between items, which describe the likelihood of encountering a new element given some prior element (Henin et al., 2021; Fox et al., 2020; comprehensively reviewed in Dehaene et al., 2015). A recent human ECoG study by Henin et al. (2021) suggests that, during the learning of sequencing relationships in continuous speech, relatively early sensory processing regions such as superior temporal gyrus (STG) instantiate representations that primarily track transition probability (or, conversely, surprisal in a local context) over time, responding rapidly to simple regularities in low-level sensory input at the level of both syllables and words. Meanwhile, the same study determined that a broad range of sites, including ventrolateral prefrontal cortex and anterior temporal regions, instantiate representations that appear to track higher-order units, but are comprised of ordinal encodings of lower-level constituents (Henin et al., 2021). Findings in other species (Shimbo et al., 2021; Carpenter et al., 2018; Ninokura et al., 2004) similarly suggest that ordinal encodings are a vital and ubiquitous component of cognitive representations of structure during sequence processing.

By triangulating such evidence, efforts to formalise the neurocomputational definitions of cognitive representations during sequence processing have become increasingly sophisticated (see Chapter 4). In line with the findings described above, our own computational model (Calmus et al., 2019) posits that bottom-up, stimulus-driven activity and top-down contextual signals, driven by prior learning, organise continuous input into discrete elements, which can then be manipulated using stereotyped processes, perhaps even akin to parsing algorithms, that produce ordinally tagged, combinatorial representations encompassing flat or hierarchical structure. Specifically, we suggest

that detection of ordinal serial position involves dorsolateral prefrontal cortex (DLPFC), motor and premotor cortex, although relevant encodings also appear to be present within the hippocampus (see Chapter 4). Meanwhile, we propose that ventrolateral prefrontal cortex is involved in sequence salience filtering, movement of items, and working memory maintenance of dependency representations in a linear or tree-like structure using an ordinal code. However, our ability to falsify models of this magnitude is limited by a scarcity of interpretable neural evidence for the configuration and dynamics of the cognitive representations involved. Indeed, our model makes relatively few pronouncements about interactions at the whole-brain level, avoiding in particular a specification of the precise mechanisms by which combinatorial encodings might arise from or interact with representations of transition probabilities. We consider that the apparent representation of transition probabilities in lower sensory regions is consistent with the predictive coding framework (Summerfield & de Lange, 2014; Friston & Kiebel, 2009; Rao & Ballard, 1999), which suggests that context-driven predictions of input from higher-order regions interact with input from lower-order regions to produce prediction errors. Specifically, if higher order regions pass down sequencing predictions to lower level cortical regions, the surprisal response to new elements in lower-level regions will covary with the transition probability of those elements. However, we again require improved methods for describing time-resolved neural dynamics in order to provide further evidence for these hypotheses. We also require neural data recorded at the appropriate spatiotemporal resolution to capture these processes (for example electrocorticography, or ECoG, which can reveal neural responses unfolding on the order of milliseconds and at a spatial resolution of 5-10 millimetres (Asano et al., 2005).

Traditional univariate analyses on electrophysiological data (see Chapter 3) are essential to understand the raw responses associated with specific stimulus features, and to establish correspondence with existing studies. However, these traditional contrastive approaches have shortcomings, including an overriding focus on the brain's electrical properties rather than its representational properties, when, as in this case, cognitive representations are increasingly the researcher's true interest. Additionally, many neuronal codes appear to be spatially distributed (as discussed in Chapter 4; and as for example recently demonstrated in auditory cortex by Fox et al., 2020). Studies using local field potentials or other meso-scale activity measures have frequently found evidence for high dimensional, task-relevant neural representations, accessible only when recordings

from multiple contacts or regions are synergistically aggregated for analysis (Tsuchiya et al., 2008; Haxby et al., 2001; Cox & Savoy, 2003). Consequently, valuable information is discarded or lost when aggregating across contacts or sites in univariate electrophysiological analyses.

Multivariate pattern analysis methods (also known as also known as multi-voxel pattern analysis) are a proven alternative to traditional univariate contrasts. These methods include pattern classification (or *decoding*) analyses, and *representational similarity analysis* (RSA; Kriegeskorte et al., 2008). All are, broadly speaking, methods that seek to aggregate neural information across sites and/or times in order to better reveal signals that correspond to mental states or cognitive representations instantiated by the subject (Diedrichsen & Kriegeskorte, 2017; Kriegeskorte et al., 2008; Haynes & Rees, 2006; Norman et al., 2006). They do this by establishing the *representational geometry* across a set of sites; that is, the way in which neural responses vary between conditions or trials at a population-coding level. Decoding methods encompass techniques that utilise training and testing of classifiers on multiple trials to characterise the representational geometry (typically using simple classifiers such as support vector machines, but also artificial neural networks including deep neural networks; see Kuntzelman et al., 2021). RSA, however, characterises the representational geometry at the level of either single-trials or conditions by taking simple pairwise distance measurements between the high-dimensional multivariate patterns under different conditions, with no training or testing required. Although the two methods are based on contrasting assumptions about the underlying data (Kriegeskorte & Douglas, 2019), RSA is applicable in many cases where a decoding analysis would be appropriate, with the benefit of reduced computational requirements. However, RSA typically does not yield an interpretable view of dynamics at a high temporal resolution. Whilst RSA conducted on time-windowed information is an established technique (see, for example, Chen et al., 2016), only a single publication to date has demonstrated the use of RSA to generate and visualise very high-resolution time-resolved representational signatures (Lin et al., 2019). Given that the publication in question analysed single-electrode recordings in rhesus macaques during a serial visual presentation task, no such analyses have been published on data in humans, under a sequence learning task, or using ECoG. We sought to employ longstanding, recent and novel multivariate analysis methods, including a high-resolution, time-resolved approach, to obtain insights into representational dynamics during the presentation of auditory sequences.

To investigate the dynamics of sequence processing, we collected ECoG data in human clinical patients undertaking an auditory artificial grammar learning (AGL) paradigm, in which sequences of syllables were presented that either fitted or violated a pre-learned pattern. We then processed this recorded data using non-traditional multivariate analyses in order to reveal non-directed functional connectivity between regions; similarity between responses to syllables across time (after Henin et al., 2021); time-resolved representational dissimilarity (building on Lin et al., 2019); and directed inter-regional flow of responsiveness to specific dimensions of the stimuli. The latter, novel method we call *representational Granger causality*.

Our findings showed that region-region multivariate ECoG similarity under the task was reflective of known functional connectivity patterns recorded using other imaging techniques. We also found, in line with the results of Henin et al. (2021) and our model, that ordinality is potentially an organising feature of a number of regional encodings elicited by the AGL task, particularly those arising in left precentral gyrus, inferior frontal gyrus pars triangularis, and supramarginal gyrus. Time-resolved dissimilarity revealed regional sensitivity to sequential stimuli across the language network during and following presentation, forming dynamic motifs of varying complexity. Causal analysis of these dynamics revealed significant outflow of item information from sensory cortex to other areas of the language network including superior temporal gyrus and inferior frontal gyrus, as well as potential signatures of feedback signals, seemingly encoding expectation violation, from multiple sources including precentral gyrus and angular gyrus. Our findings finally suggested that dynamic multivariate methods, including representational Granger causality, have considerable promise as tools for establishing the neurocomputational foundations of cognition.

5.3 Methods

5.3.1 Participant recruitment and ethics

Participant recruitment is described in 3.3.1. The same participants contributed to this analysis. All (n = 12) participants were adult neurosurgical patients (7 male, 5 female, ages 19–55, median 33 years) diagnosed with medically refractory epilepsy and undergoing chronic intracranial pre-surgical monitoring at University of Iowa Hospitals and Clinics (UIHC) to identify seizure foci suitable for resection. Research protocols were approved by the University of Iowa Institutional Review Board (IRB ID No.: 200112047) and National Institutes of Health. Participation in the study did not impact on clinical monitoring or management. Informed written consent was obtained from each subject prior to their participation in the study, and participants were free to withdraw consent at any time without any impact on clinical evaluation. Sessions were suspended for at least three hours if a seizure occurred, and only resumed if the participant was willing and alert. These ethical guidelines are consistent with European Ethical Guidelines (Helsinki Declaration and H2020 EU guidelines).

5.3.2 Task design

The task was an artificial grammar learning task, the *AxB* task, previously described in detail in 3.3.2. The analysis here uses the same raw behavioural and electrophysiological data analysed in Chapter 3. Participants were exposed to auditory sequences fitting an **XAB** or **AXB** pattern, where **X** was highly variable and uninformative, there were 2 salient **A** and **B** items, and **A** was always paired with a specific **B**. Participants were then tested on their ability to judge the grammaticality of novel sequences consistent with, or violating, this relationship.

5.3.3 Electrode configuration and acquisition

As described in 3.3.3, participants were implanted with clinically indicated subdural ECoG arrays and depth electrodes. Local Field Potential (LFP) data was acquired with a Neuralynx Atlas system (Neuralynx, Bozeman, MT, USA), amplified, hardware bandpass filtered to between 0.1 Hz and 500 Hz (5 dB/octave roll-off), and digitised at a sample rate of 2000 Hz. A room microphone and auxiliary sound output were both also simultaneously recorded using the Neuralynx system, along with digital events generated by the task script. Behavioural data for each participant was also stored by the task script for offline analysis. Electrode locations were confirmed in each patient by

the clinical and HBRL team via co-registration of pre- and post-implantation structural imaging supported by intraoperative photography (Nourski & Howard, 2015). Recording sites were linearly co-registered to the MNI152 T1 standard brain, placing them into MNI coordinate space, and subsequently assigned to standard anatomical regions of interest (ROIs) via anatomical reconstruction of electrode location. This was aided in the case of subdural arrays by automated parcellation of cortical gyri (Destrieux et al., 2010, 2017) using the FreeSurfer image analysis software suite (http://surfer.nmr.mgh.harvard.edu/), and in the case of depth electrodes with reference to structural MRI sections. Electrode recordings from sites implicated in seizure activity were excluded from subsequent electrophysiological analysis, and analysis was restricted to recording sites assigned to the ROIs listed in Table 3.2 (see Chapter 3).

5.3.4 Pre-processing of channel recordings

Electrophysiological data were analysed offline using a custom MATLAB library. Neuralynx data were first imported into MATLAB, and ECoG/LFP data and line noise removed using the demodulated band transform (DBT; see Kovach & Gander, 2016). Data was down-sampled to 500 Hz and further denoised by discarding the first principal component of the singular value decomposition (SVD) of the highpass-filtered signal (cutoff 160 Hz; see also methods of Kumar et al., 2020). To minimise jitter, recorded digital events were aligned with the auxiliary sound input (down to the per-sample level at 2 kHz) by shifting them using a custom event alignment script. Any data fragmented across multiple recording files was concatenated into one session, and continuous LFP data from selected channels epoched from -1500 to 4000 milliseconds around stimulus onset. Subsequently, epoched data was grouped into trials falling into each of the experimental conditions (chiefly: "consistent", "violation", "adjacent/**XAB**", "nonadjacent/**AXB**").

5.3.5 General computation of representational dissimilarity

Following pre-processing, a variety of multivariate analyses were conducted on the neural data. Each varied in their purpose and design, but all shared a common trait of operating on *representational dissimilarity matrices* (RDMs). Here, the process for deriving an RDM is explained. To abstract away from the activity patterns present in the raw data, a matrix of pairwise comparisons, the RDM, is computed between neural (or behavioural, or simulated) responses under each experimental condition. An RDM is thus comprised of many pairwise dissimilarity measures, each entry characterising the distance between responses to two stimuli or conditions. For example, if each

electrophysiological response to be analysed forms an activity pattern comprising 1000 samples, each pairwise dissimilarity value of the RDM will be a scalar distance computed across the 1000 samples of the pattern, between the two patterns. Any one of a host of distance measures, *d*, can be used for this purpose, for example correlation distance:

$$d(p, q) = 1 - \rho_{p,q}$$

where ρ is the correlation coefficient (Spearman or Pearson *r*) between two vectors *p* and *q*, the activity patterns. Alternatively, one can compute Euclidean distance, as follows:

$$d(p, q) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

where *p* and *q* are likewise two activity patterns to be compared. Any spatiotemporal response vectors can serve as the activity patterns from which an RDM is derived, including any neural or behavioural measurement, or computational model. By this method, data of any modality can be summarised in an RDM, permitting abstract representational characteristics to be compared and contrasted at a higher level.

Here, RDMs were computed across multiple chosen contacts within a specific ROI by taking the mean response across selected trials within a given time window for every chosen contact, concatenating the resulting per-contact patterns into a single vector per condition, and using these per-condition signals as the inputs to the pairwise distance measure. In this way, it was possible to produce an RDM summarising the representational dissimilarity patterns during a particular time window for a given *a priori* defined anatomical region of interest.

5.3.6 Neural pattern similarity

One form of analysis that can be conducted using RDMs is the representational similarity analysis (RSA, Kriegeskorte et al., 2008). RSA is a multivariate analytical technique that permits the comparison of neural data from one source with data from other modalities, species or brain regions, or against computational, behavioural or simple categorical models. The general principle of comparison is summarised in Figure 5.1 (see also Kriegeskorte et al., 2008). In canonical RSA, previously calculated RDMs (A and B, for example representing the true data and a computational model, respectively) are compared by first linearising the upper or lower triangle of the symmetric

matrix, and undertaking a correlation between the two resultant vectors. By shuffling one vector repeatedly (C), a permutation test can generate percentile thresholds for significance testing from a null distribution of *r* values (D). Thus, RDMs can be compared and the significance of that comparison tested. The final *r* value reflects the degree of second-order isomorphism between the two RDM sources; that is, the degree to which the responses in RDMs A and B vary in a similar manner between conditions.

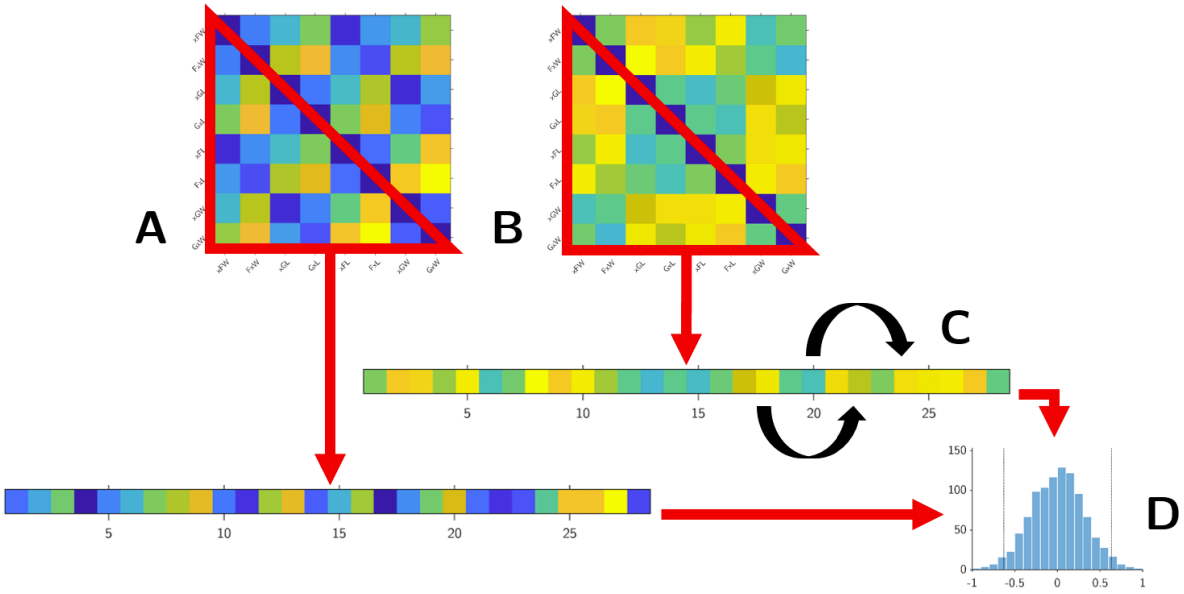


Figure 5.1: **Canonical method for testing RDM correlations (representational similarity analysis).** RDMs A and B are shaped into vectors and correlated with each other. Correlations may be tested by a permutation analysis if necessary, by shuffling one vector (C) to characterise the null distribution of correlation coefficients (D).

A *neural pattern similarity* (NPS) analysis involves conducting RSA on multiple sources of neural data and comparing them representationally, namely comparing representations in one region with those of another. In this study, NPS was conducted for two reasons: in order to determine whether the chosen regions of interest (ROI) contained meaningful representational content, as a precursor to further analyses; and secondly to broadly characterise regional functional divisions under the AxB task in this cohort.

To calculate neural pattern similarity, RSA was conducted between regions for each participant in the study, and results were then summarised across the cohort. RDMs were calculated on time-domain neural data for each of the n = 12 implanted participants (see Chapter 3), using a custom MATLAB library. The RDMs were derived from correlation distances between ECoG signals

recorded during the grammaticality testing phase. Pairwise Pearson correlation distances were obtained by comparing spatiotemporal vectors under every pairing of the 8 experimental conditions (the 8 conditions: $\mathbf{xA_1B_1}$, $\mathbf{A_1xB_1}$, $\mathbf{xA_2B_2}$, $\mathbf{A_2xB_2}$, $\mathbf{xA_1B_2}$, $\mathbf{A_1xB_2}$, $\mathbf{xA_2B_1}$ and $\mathbf{A_2xB_1}$). Each spatiotemporal vector contained time-domain data from a portion of every epoch, concatenated (comprising the ECoG signal recorded between -1.5 and +4 seconds relative to stimulus onset, for every trial), and from all contacts within an *a priori* defined ROI from the table listed in 3.3.3. Pairwise distance computations produced, for each participant, a single 8 × 8 RDM for every ROI that took into account neural response variability at the single-trial level.

Subsequently, pairs of RDMs (each representing an ROI) were themselves correlated with each other by following a variation of the procedure described in Figure 5.1. To correlate RDMs, the Spearman rank correlation coefficient was used; rank-based comparison methods are recommended here to avoid assumptions of a linear match between RDMs, since the activity patterns in each RDM can exhibit different noise characteristics (Kriegeskorte et al., 2008). This produced a matrix of region-region *r* values which were averaged across the cohort (excluding pairs of regions absent in a given participant), and masked to $\alpha = .05$ by performing a one-tailed, one-sample *t*-test between zero and the group's *r* values for the comparison. This departs from the canonical single-subject RSA procedure described above, but can be justified by the fact that *r* values on random datasets occupy a finite-bounded approximation of the normal distribution (Hotelling, 1953). The result of this process was a significance-masked matrix of similarities between the representations instantiated within different ROIs across the cohort. ROI-ROI comparisons with higher *r* values can be interpreted as being more similar in their representational geometry, and lower values less similar.

To aid interpretation of this matrix of correlation coefficients, the unmasked Spearman *r* data was transformed into a dendrogram, or hierarchical tree diagram, in which leaves represent brain regions and the branch size between two leaves is proportional to the distance between them. Distances were computed from the average correlation coefficients across the cohort, excluding region-region pairings absent in each participant, using the simple formula $d = 1 - r$, where *d* and *r* are the distances and correlation coefficients, respectively. The dendrogram was computed by agglomerative hierarchical clustering (for an efficient method, see Day & Edelsbrunner, 1984). Using this algorithm, items are iteratively merged into clusters depending on their distances from one another, beginning with the closest items. As clusters are formed, the closest clusters or items are

likewise merged, until all items have been merged. This requires a *linkage function* that defines how to compute the distance between two clusters. Here, agglomerative hierarchical clustering was conducted using an un-weighted average linkage function, which defines the distance between two clusters as the arithmetic mean of the distances between all possible pairings of items across the two clusters. During this clustering process, to mitigate any artefactual clustering caused by the distribution of coverage across the cohort rather than neural pattern similarity, missing values were completed by mean imputation. Simply put, this meant that region-region distances missing in any individual (for example between hemispheres, where a participant typically only had coverage on one side or the other) were replaced by the mean distance across the entire correlation coefficient matrix for that individual. This yielded a tree where similarly responding regions clustered together and were hierarchically grouped by their single-trial response patterns.

5.3.7 Dissimilarity between responses to syllables across time

A second analysis was undertaken to determine, for each region, the similarities to different classes of syllable over time, inspired by Henin et al. (2021). Recorded epochs were again separated into 8 experimental conditions (every combination of $\mathbf{xA_1B_1}$, $\mathbf{A_1xB_1}$, $\mathbf{xA_2B_2}$, $\mathbf{A_2xB_2}$, $\mathbf{xA_1B_2}$, $\mathbf{A_1xB_2}$, $\mathbf{xA_2B_1}$ and $\mathbf{A_2xB_1}$). Data was averaged across the 12 trials under each condition, and then windowed in time to capture the first syllable and its subsequent inter-stimulus interval (0 to 750 ms), the second syllable and subsequent inter-stimulus interval (750 to 1500 ms), and third syllable over the same period (1500 to 2250 ms). This produced a single 24 × 24 RDM containing responses to 3 syllables under 8 conditions, for each region, describing the dissimilarities between responses to each class of syllable at each time position (first or second position for **X** and **A** items; always third for **B** items). This data was then subjected to metric multidimensional scaling (Torgerson, 1952, 1958; Gower, 1966; described in more detail in the next subsection) to produce visualisations of the syllables, drawn as points in 2-dimensional Cartesian space. Points were visually grouped by drawing 95% confidence ellipses drawn around specific categories of stimulus (for example, all syllables of the same ordinal position). Ellipses were positioned at the centroid of the category of interest, aligned with the largest eigenvector of the point distribution, proportioned to encapsulate the points using the standard deviation along the principal eigenvectors, and scaled to form confidence bounds according to the cumulative Chi-square distribution (2 degrees of freedom) at the 95th percentile.

5.3.8 Dynamic representational dissimilarity

The neural pattern similarity analysis was in principle sensitive to second-order isomorphism between regions, when considering responses under all trials of the task, and the syllable-level analysis was sensitive to representational differences between specific time windows. However, a time-resolved analysis was required to clarify the dynamic representational characteristics of each region at a high temporal resolution. One possible method by which to characterise representational dissimilarity over time is to effectively produce many "frame-by-frame" RDMs, each associated with a snapshot in time (Lin et al., 2019). As before, data is viewed not on the level of individual electrode channels, but rather as spatiotemporal vectors derived from multiple channels in a given region of interest (ROI), across multiple time points in a given window of interest. In this case, however, the window of interest is moved across the input channel data to form different spatiotemporal vectors at each point in time. This approach is familiar to dynamical systems theorists, where it is known as "time-delay embedding" or "state space embedding" (Braaksma et al., 1985; Whitney, 1936), and it has more recently been employed in neuroscience (for example Lin et al., 2019; Anderson et al., 2006). As when producing a "static" RDM, the dissimilarity between two conditions is the distance between two corresponding spatiotemporal vectors. However, by forming time-varying spatiotemporal vectors, the process of time-delay embedding produces multiple RDMs over time (or, equivalently, a single three-dimensional RDM containing time-varying distance information).

To maximise the reliability of these repeated distance measurements, data was first spatially pre-whitened (i.e., subjected to multivariate noise normalisation, see Walther et al., 2016). This mitigates the impact of noise correlated between recording sites, a problem affecting many imaging modalities including intracranial recordings, using the estimated variance-covariance matrix of the recorded data across trials to de-correlate the noise between sites (Diedrichsen & Kriegeskorte, 2017). To pre-whiten data, zero-phase components analysis (ZCA; see Bell & Sejnowski, 1997) was used to estimate a whitening matrix from a pre-stimulus baseline period (-1 to -0.2 seconds relative to stimulus sequence onset). Recorded per-trial intracranial activity was then de-meaned, and the spatially pre-whitened data taken to be the matrix product of this de-meaned data and the whitening matrix. Note that whilst the singular value decomposition (SVD) denoising process described previously (see "Pre-processing" above) spatially filters continuous channel data to minimise correlations across contacts, the process here differs firstly by explicitly de-correlating, rather than

outright removing, correlated components; and secondly by making reference to the per-trial baseline to distinguish background activity from possible task-relevant activity, rather than acting on correlated components detected solely from continuous data.

The steps of the time-resolved representational analysis are illustrated in Figure 5.2. Trial-averaged event-related potentials (ERPs) were first computed for each channel in an ROI under a given condition, and time-windowed ERP data for every channel concatenated into a spatiotemporal vector (Figure 5.2A). Spatiotemporal vectors were thus computed for each of the ROIs under each of the experimental conditions, and spanning specific windows in time. The same computations were conducted for every time-point in the trial by iteratively sliding the temporal window (a 300 ms rectangular window, sliding by 5 ms per iteration, producing a time series of spatiotemporal vectors with an effective sampling rate of 200 Hz). Although epochs in the study strictly ran from -1500 to 4000 ms post-stimulus onset, only data between 0 ms and 3000 ms (more specifically, from windows centred at these times) was considered in the DRD analysis, to omit the subsequent period containing participants' motor responses.

For a given ROI, within a given time window, pairwise comparisons were then conducted between all conditions to produce a representational dissimilarity matrix (RDM; Figure 5.2B). Pairs of spatially pre-whitened spatiotemporal vectors were compared by computing the Euclidean distance between them. More precisely, because channel data was pre-whitened prior to analysis, the entry in each cell of the RDM was the Mahalanobis distance between conditions (since Euclidean distances on pre-whitened data are equivalent to Mahalanobis distances; see Diedrichsen & Kriegeskorte, 2017). This produced 8 × 8 distance measurements, forming an RDM symmetrical about the diagonal. For each time window, an 8 × 8 RDM was thus produced from trial-averaged data under each of the factorial task conditions.

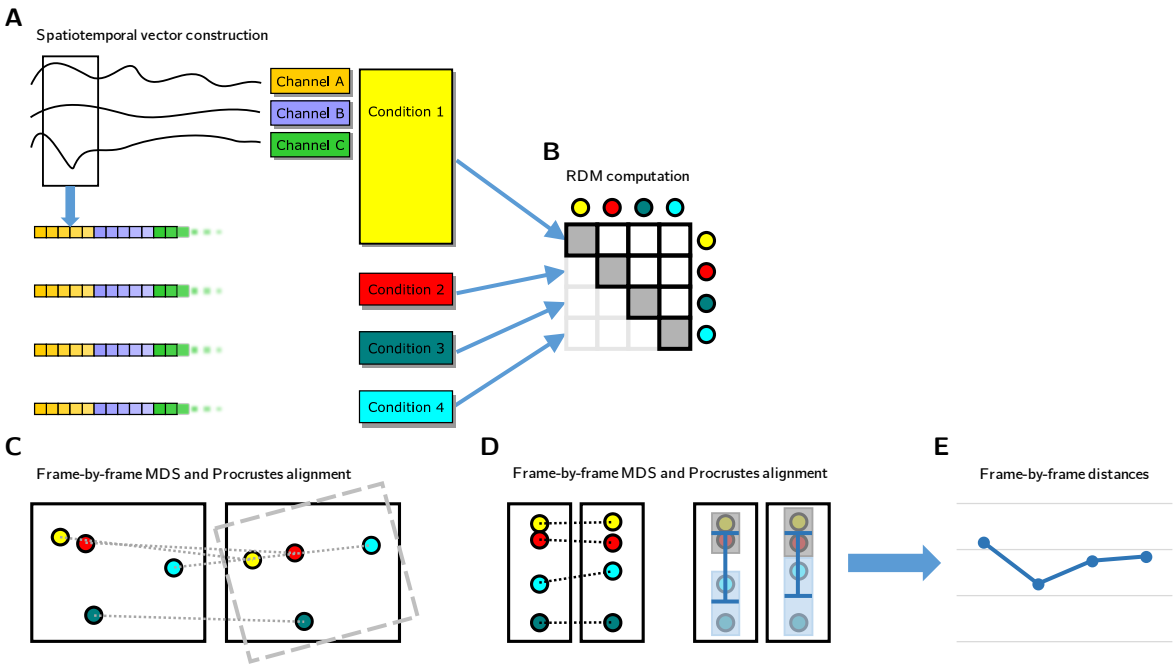


Figure 5.2: **Time-resolved representational analysis, including Procrustes multidimensional scaling (pMDS).** See main text for further details. A) Spatiotemporal vectors are produced by generating time-delay embeddings from pre-whitened trial-averaged data at multiple sites and times. Here, embeddings were produced by a simple sliding window 300 ms in duration, iteratively shifted by 5 ms. B) Distances between spatiotemporal vectors are entered into a representational dissimilarity matrix (RDM), producing one RDM per window (i.e. an RDM every 5 ms). C) RDMs can be visualised in low-dimensional space – shown for clarity in 2D – by multidimensional scaling (MDS), and matched across time-points by using the Procrustes algorithm, which scales and translates points to optimally align them, in a process known as Procrustes multidimensional scaling. D) Undertaken in one-dimensional space, pMDS over time produces a set of waveforms showing changes in representational similarity between conditions over time. To group points representing multiple experimental conditions, we can find their centroid using the arithmetic mean. (E) The distance between two centroids in low dimensional space produces what we call a *dynamic representational dissimilarity* (DRD) wave useful for downstream analysis.

For each time window (each frame of 300ms, slid by 5ms), metric multidimensional scaling (MDS) was then conducted on the resulting RDM. Metric MDS (also known as classical MDS; Torgerson, 1952, 1958; Gower, 1966) is a method of dimensionality reduction that transforms pairwise dissimilarities between objects, as stored in a distance matrix, into a visualisation of the objects as points in an abstract, low-dimensional Cartesian space, whilst attempting to preserve the relative dissimilarities between them. It is rarely possible to exactly reproduce the original distances in low-dimensional space using MDS, but the approach produces a set of low-dimensional points where the original distances and low-dimensional distances are maximally linearly correlated. The

method is closely related to principal component analysis (PCA; Pearson, 1901; Hotelling, 1933), and the solution for a given distance matrix is therefore unique and deterministically reproducible.

The left box of Figure 5.2C illustrates this process with an output of 2 dimensions. The representations under each condition are approximately as proportionally close to each other in 2D as the original spatiotemporal vectors. However, MDS on different distance matrices will not necessarily yield points that are rotated or translated consistently in low-dimensional space. As a result, it is not typical to produce multiple MDS results over time, as these cannot by default be readily compared with each other. Historically, this has made it difficult to visualise representational dynamics. Here, however, useful time-aligned MDS results were produced over time by adapting a method employed by Lin et al in their 2019 conference paper. All MDS results were transformed using a generalised Procrustes analysis (Gower, 1975) to produce a globally optimal alignment between all MDS frames (Figure 5.2C, left versus right boxes; see Lin et al, 2019). The generalised Procrustes algorithm is a method for optimally aligning two or more sets of points by rotating, translating and scaling them to match as closely as possible, minimising a goodness-of-fit criterion. As in Lin et al. (2019), scaling was explicitly disallowed during alignment, since it would invalidate the meaning of the magnitude of distances over time, but rotation and translation were permitted.

Procrustes-aligned multidimensional scaling (pMDS) results can be plotted in arbitrary low dimensions, including as two-dimensional plots over time (i.e. three-dimensional visualisations). However, the simplest possible space that can be visualised is a 1-dimensional representational embedding over time (i.e. a two dimensional visualisation, for example with time on the abscissa and representational space on the ordinate axis). This is potentially the most readily interpretable and intuitive of the visualisations possible with pMDS, with representations separating vertically when they are different, and converging when they are similar (see Figure 5.2D, left boxes). As with all of the measures computed in this analysis, it was possible to produce a pMDS visualisation for a specific ROI, or a collection of regions or contacts, or on a whole-brain basis.

In this study, the pMDS visualisation was used primarily as an interim representation in a more involved analysis, which we call a *dynamic representational dissimilarity* analysis. Specifically, pMDS was first used to produce a 2D mapping of representational space over time, but the centroids of multiple related conditions were then determined by arithmetic mean (for example, the low-

dimensional embeddings of all "grammatical" conditions were averaged together, as were those of all "ungrammatical" conditions, to compare "grammatical" and "ungrammatical" representations; likewise for $\mathbf{A_1}$ versus $\mathbf{A_2}$, and $\mathbf{B_1}$ versus $\mathbf{B_2}$; see Figure 5.2D, right boxes). For most representational contrasts (i.e., comparisons on the basis of grammaticality, **A** or **B**) this yielded two average traces in low-dimensional space, between which a simple difference wave was computed, to produce a time series of distances between each set of conditions over time, which we call a dynamic representational dissimilarity plot (or DRD plot; Figure 5.2E). The greater the overall difference between the grouped conditions at a specific point in time, the greater the magnitude of the DRD wave. The DRD waves could therefore be conceived of as multivariate contrasts or, equivalently, as a measure of the encoding strength of a given feature or representation over time. This same method was extended to allow measurement of dissimilarity between more than two conditions (for example in the case of first or second syllable identity, which unlike the **B** item could take three categorical values: $\mathbf{A_1}$, **X** or $\mathbf{A_2}$). This was accomplished by computing the pairwise centroidal distances between all contrasted groups and taking the DRD to be the maximum of these at a given time. We refer to these syllabic identity contrasts as $\mathbf{S_1}$ (for items in the first position) and $\mathbf{S_2}$ (for items in the second position). $\mathbf{S_3}$, the syllabic identity in the third position, was by definition identical to **B**.

At any time point, this dynamic representational dissimilarity was non-trivially related to the original distance matrix, and could not be determined simply by averaging over the inter-group (e.g. all grammatical to ungrammatical) distances. Rather, because classical MDS is based on eigenvalue decomposition and closely related to PCA, and because we expressly produced a one-dimensional embedding at each time point, the DRD value at a given time related to the first principal component of the representational dissimilarity matrix; that is, the high-dimensional axis along which the RDM exhibited maximum variance. Representational dissimilarities are known to be positively biased when estimated on noisy data (Walther et al., 2016), but PCA-based methods have been widely used for denoising purposes, by dropping the smallest principal components (those that explain the least variance) from data. Therefore, one-dimensional MDS produces output of a visually appropriate dimensionality, and also has the potential advantage of reducing noise in the final distance estimates.

Cohort-level dynamic representational dissimilarity was computed by producing pMDS embeddings for each ROI of every participant and then, by following the method above,

transforming each pMDS embedding into multiple DRD waves, one per contrast. These included **A** identity ($\mathbf{A_1}$ versus $\mathbf{A_2}$); **B** identity ($\mathbf{B_1}$ versus $\mathbf{B_2}$); grammaticality (grammatical versus ungrammatical); 1st word identity, $\mathbf{S_1}$ (i.e. max[d($\mathbf{A_1}$, $\mathbf{A_2}$), d($\mathbf{A_1}$,**X**), d($\mathbf{A_2}$, **X**)], where d was the Euclidean distance function, and item labels were obtained from the first position of the entire sequence); and 2nd word identity, $\mathbf{S_2}$ (likewise, for labels varying in the second position). For visualisation only, DRD waves were subsequently normalised to between 0 and 1, to account for differing signal-to-noise ratios across participants, expected with differing electrode placements and recording conditions. Normalising enabled meaningful aggregation over multiple subjects, to produce an average DRD wave for each ROI, with limits, for each of the **A**, **B**, $\mathbf{S_1}$, $\mathbf{S_2}$, and "grammaticality" contrasts across the cohort. The resulting plots depicted the contour of representational dissimilarities between conditions in a given region over time, across the cohort.

5.3.9 Representational Granger causality

We undertook causal analysis of multiple DRD datasets (**A**; **B**; "grammaticality"; $\mathbf{S_1}$, $\mathbf{S_2}$, and $\mathbf{S_3}$; generated as described in 5.3.8) using a specific measure of directed functional connectivity known as Granger causality. Granger causality (or "G-causality") is a mathematical notion of causality originally formalised for use on economic data (Granger, 1969; elaborating upon Wiener, 1956), but it has since been applied with great efficacy to neural datasets (Seth et al., 2015). Granger causality describes the extent to which one signal, X, can be used to predict later samples of a second signal, Y, better than signal Y can be used to predict itself. Since we were seeking a characterisation of complex interactions between multiple ROIs, we specifically relied upon the notion of *multivariate conditional Granger causality* (MVGC; first defined in Geweke, 1984, 1982). MVGC expands upon the original pairwise notion of causality by determining the influence of X on Y conditional on other time series Z (that is, $X \rightarrow Y \mid Z$). This permits the testing of causal relationships between more than two time series at once, and, given data from sufficient sites, is therefore able to adjudicate between direct and indirect causal associations (for example, for three known time series X, Y, Z, adjudicating between $X \rightarrow Y$ versus $Z \rightarrow X$, $Z \rightarrow Y$). Here, to distinguish our overall method (which uses signals derived from time-resolved representational dissimilarity) from existing methods (which determine connectivity directly from channel or voxel time series) we simply call it "representational Granger causality".

Representational Granger causality was computed between the regional DRD waves for each of the representational contrasts (**A**, **B**, "grammaticality", $\mathbf{S}_{1\ldots3}$) in MATLAB using the MVGC toolbox (Barnett & Seth, 2014). Multivariate conditional G-causality was estimated between all pairs of ROIs (except a region and itself) from the DRD signals obtained in every suitable participant (n = 12 subjects), after first down-sampling the DRD wave in the time domain by a factor of 5 for computational efficiency (producing an effective sampling rate of 40 Hz). The MVGC toolbox allows causality to be estimated from multiple "samples", each a set of concurrently obtained signals. To perform a cohort-level analysis, each subject's DRD dataset was treated as one sample, yielding 12 samples. As in many neuroscientific analyses, this assumes the existence of a stereotypical system behaving similarly in each subject.

To estimate G-causality, a model must first be fitted to the data, in this case a multivariate vector autoregression (MVAR) model, which describes the evolution of each timeseries in terms of the histories of the other available time series, thus:

$$X_t = \sum_{k=1}^{p} A_k \cdot X_{t-k} + E_t$$

where X is a multivariate time series (a matrix of measurements across time for n variables of interest); A_k is an $n \times n$ matrix of regression coefficients; E_t the n residuals, which effectively represent white noise; and p is the order of the model, which describes how many measurements prior to the present time t are incorporated into the prediction of the new measurement.

To support accurate causal estimates, the order of the MVAR model must be appropriately selected. This can be accomplished using an information criterion, an objective measure of the balance between parsimony and statistical goodness of fit. Here, MVAR models were considered up to order 40, and information criteria computed for each using the MVGC toolbox. The model with the minimum criterion was then selected. The criterion of interest was a version of the Akaike information criterion (AIC; Akaike, 1974; specifically AIC_C, featuring bias correction to improve the accuracy of estimates on small sample sizes; Hurvich & Tsai, 1989). Using AIC to select the optimal model for each DRD dataset (**A**, **B**, "grammaticality", $\mathbf{S}_{1\ldots3}$), the MVAR in every case was of order 2. The coefficients of the order-2 MVAR models were then estimated using the Levinson-Whittle-Wiggins-Robinson approach, a recursive method for fitting MVAR models (Whittle, 1963).

MVAR coefficients were then checked for undefined or infinite values to ensure they did not indicate a rank-deficient or ill-conditioned regression.

In the event that data is non-stationary, or incorporates decaying memory effects, Granger causality analysis may be inappropriate and can yield spurious results. More precisely, to ensure a valid analysis, the VAR coefficients must be stable, defining a covariance-stationary process (Barnett & Seth, 2014). VAR stability was tested using the MVGC library by assessing the spectral radius of the VAR (the largest absolute eigenvalue of the VAR coefficient matrix, A_k). Models with a spectral radius <1 can be proven to satisfy the stability condition (Barnett & Seth, 2014). In this case, by this test, all selected VAR models were stable, and analysis proceeded. An *autocovariance sequenc*e, which describes the covariance between X_t and X_{t-k} for different values of k, was then estimated from the MVAR, with time-domain pairwise-conditional G-causality computed in turn from this sequence. Finally, for each pair of regions, significance of the causal relationship between them was determined using Granger's F-test, corrected for multiple comparisons using the false discovery rate.

Following computation of the Granger causality results, a directed graph (or *digraph*) was constructed to summarise each causality network. The digraph was constructed by computing the net flow of G-causality between all pairs of brain regions, and using these to determine edge weights. Net flow "summarises the prevailing direction of causality, net of changes in overall causality strength" (Duggento et al., 2018). The exact metric used here, *significance-masked net flow*, took into account only significant values and was defined as the significance-masked causality strength in the backwards direction subtracted from the significance-masked causality strength in the forwards direction, or, as follows:

$$F_{i,j} = ((G \circ H) - (G \circ H)^T)_{i,j}$$

where G denotes the matrix of Granger causality values, H denotes the matrix of significance masking values (H = 1 if p <= .05, or 0 otherwise), "^{T}" denotes matrix transposition and "$\circ$" denotes the Hadamard product (element-wise multiplication) of two matrices. It was intuited that net flow would likely be more useful as a summary of representational G-causality results than on traditional Granger causality findings, since representational G-causality should reflect the transmission of relatively targeted information. In constructing the digraph, when net flow F was strictly positive in the forwards direction between two regions i and j, the edge weight for the edge $i \rightarrow j$ was F,

otherwise the edge weight was assigned a value of 0 (that is, no edge was recorded). Self-connectivity (any edge $i \rightarrow i$) could not occur in this graph, since G-causality was not computed for these cases.

5.4 Results

5.4.1 Neural pattern similarity

The neural pattern similarity analysis resulted in a symmetric matrix of mean correlation coefficients denoting the average similarity across the entire duration of the epoch, across the cohort, between the RDMs of any two ROIs. The matrix is shown in Figure 5.3, significance masked to $\alpha = .05$. Excluding self-similarities (where $r = 1$), the range of significant mean coefficients spanned 0.10 – 0.75. It is notable that the most similar ROI pairings were not restricted to sensory cortex, but rather promised to shed light on the function of more abstruse regions, especially inferior frontal gyrus (IFG). For example, the three strongest similarities in the left hemisphere were between middle frontal gyrus (MFG) and IFG pars triangularis ($r = 0.70$); superior temporal sulcus (STS) and middle temporal gyrus (MTG; $r = 0.70$); and IFG pars triangularis and orbital gyrus ($r = 0.66$). In the right hemisphere, the three strongest similarities were between Heschl's gyrus (HG) and STS ($r = 0.75$); IFG pars triangularis and IFG pars orbitalis ($r = 0.71$); and MFG and IFG pars triangularis ($r = 0.71$).

The dendrogram of the unmasked correlation data, produced by agglomerative hierarchical clustering of distances (where distance $d = 1 - r$), is a potentially more interpretable visualisation and is shown in Figure 5.4. It can immediately be seen that many ROIs clustered on the basis of the hemisphere in which they could be found. The functional properties of the left and right hemispheres may have influenced this result, but the effect of coverage likely predominated here, despite efforts at mitigation (see 5.3.6). However, we were also able to identify key groupings of regions common to both the left and right hemispheres. This included relatively close similarities, in both hemispheres, between the responses of hippocampus and parahippocampal gyrus (PHG; both in grey); angular gyrus and supramarginal gyrus (SMG; both in blue); superior temporal gyrus (STG) and superior temporal sulcus (STS); and middle frontal gyrus (MFG) and orbital gyrus.

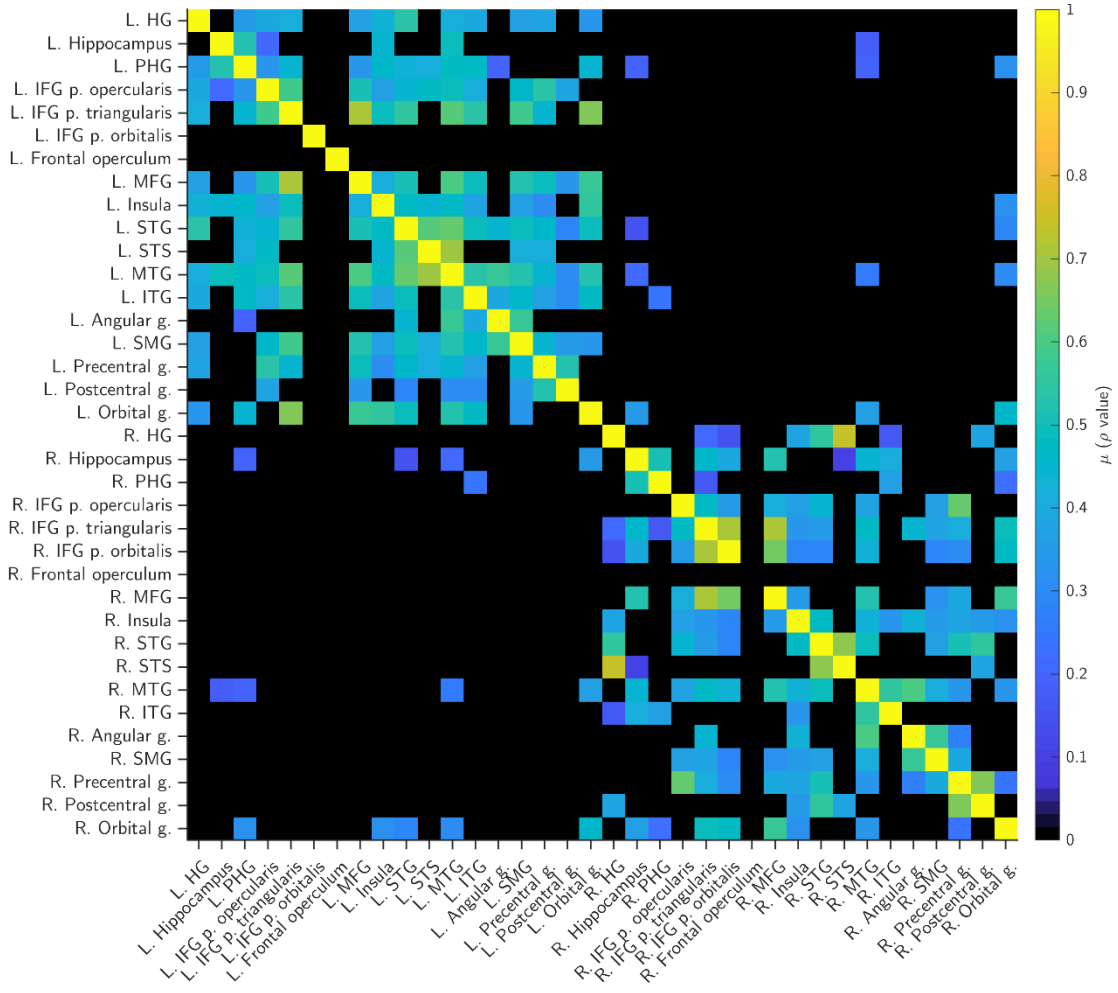


Figure 5.3: **Significant neural pattern similarity between regions across cohort (n = 12).** Mean Spearman ρ between representational dissimilarity matrices (RDMs) shown for every region of interest, averaged across the cohort, masked by significance (p = .05). RDMs were computed from trial-trial Euclidean distances. If a given pairing of regions was never present across the cohort, such comparisons were also masked out.

Some clustering was also observed between regions that are anatomically distal but known to be structurally and functionally connected, such as subregions of inferior frontal gyrus (IFG, in green) and the temporoparietal junction (TPJ, in blue). For example, left SMG more closely clustered with left IFG pars triangularis (via a linkage distance of 0.56) than to the physically nearer Heschl's gyrus (via a linkage distance of 0.61), a relationship that accorded with the original correlation matrix (left SMG to IFG p. triangularis: r = 0.58; to HG: r = 0.37). The IFG is known to communicate with the inferior parietal lobule via major tracts including the arcuate fasciculus and superior longitudinal fasciculus, which are both critical for language in humans and have homologues in the non-linguistic precursor circuits of nonhuman primates (Balezeau et al., 2020; Petrides & Pandya, 2009).

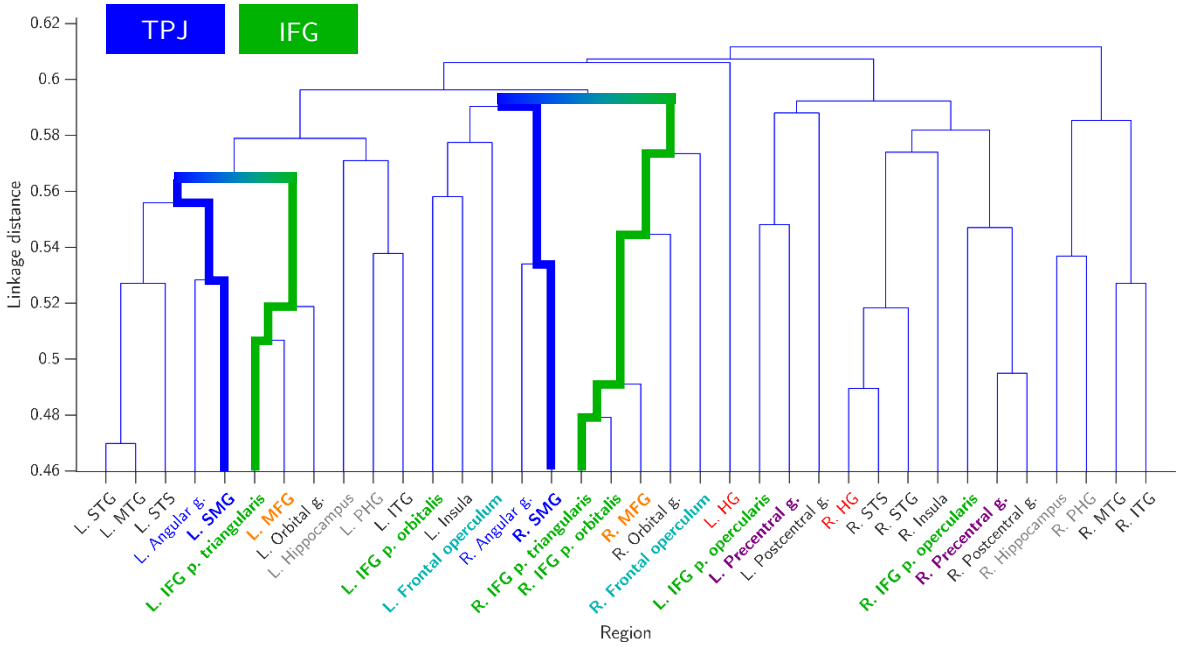


Figure 5.4: **Dendrogram showing agglomerative hierarchical clustering of neural pattern similarities across the cohort.** Clustering was based on comparing un-weighted average distances derived from the unmasked Spearman ρ matrix. In a dendrogram, the total linkage distance between clusters is displayed on the ordinate axis. The relatively strong clustering of physically distant TPJ and IFG subregions is highlighted. Selected other regions are also coloured for reference (see main text).

Conversely, adjacent subregions of IFG (all shown in green) largely clustered more strongly with non-IFG regions than they did with each other, apparently highlighting the functional heterogeneity of this area. In both hemispheres, IFG pars triangularis and middle frontal gyrus (MFG, in orange) were closely associated; as were IFG pars opercularis and precentral gyrus (in purple); and IFG pars orbitalis and frontal operculum (in cyan). These groupings accord well with the results of prior studies on the functional connectivity of IFG (for example, as determined by fMRI regional co-activation, reported in Wang et al., 2020).

Given numerous indicators that the spatiotemporal vectors of intracranial recordings contained meaningful information on the functional properties of the brain under the task, time-windowed analyses were performed as described in 5.3.7 and 5.3.8, and are reported below.

5.4.2 Dissimilarity between responses to syllables across time

In section 5.2, we revisited our earlier hypotheses (see Chapter 4) that specific regions are involved in the ordinal coding of sequences, namely dorsolateral prefrontal, motor and premotor cortex; and that ventrolateral prefrontal cortex is engaged during working memory for dependencies

using a linear or tree-like structure that likewise incorporates an ordinal code. Here, we predicted that these regions would exhibit MDS results consistent with ordinal positional coding, in contrast to other regions. Schematics are shown in Figure 5.5, depicting the approximate expected configurations of the confidence ellipses in the MDS analysis for different hypotheses. Specifically, Figure 5.5 panel A shows what we expected MDS results to look like in explicitly ordinally coding regions, by which we mean any region that either codes for ordinal position or else incorporates ordinality as part of a relational code. Conversely, panel B shows expected results in a region entirely insensitive to order. Finally, panel C acknowledges that some separation of confidence ellipses might also occur as a result of repetition suppression effects, since a change in the magnitude of a response over time should produce a characteristic pattern of Euclidean distances within and across syllables.

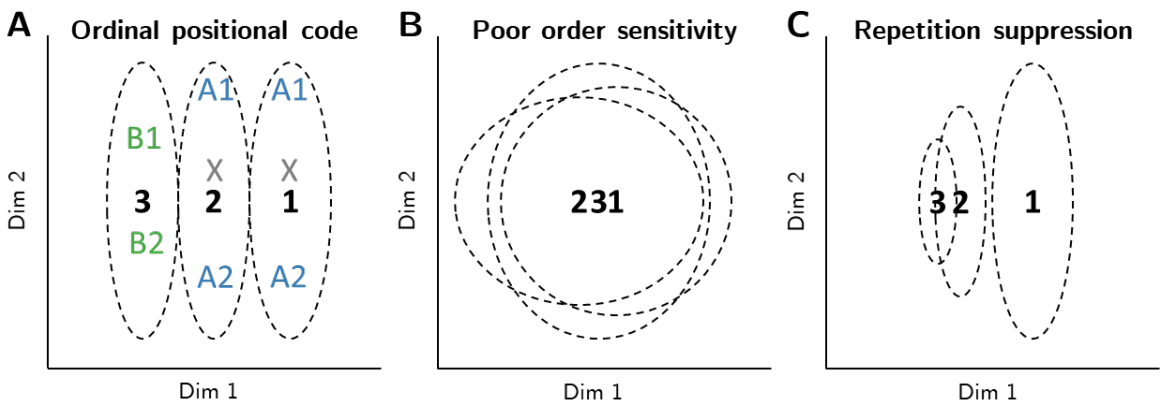


Figure 5.5: **Schematics showing predicted MDS layouts under alternative hypotheses.** The dimensions themselves are arbitrary and unitless, and only relative locations of datapoints or ellipses are therefore informative. A) Clear separation of responses into 3 distinct clusters based on item position. This is compatible with the notion of an explicit ordinal positional code. In particular, under such a coding scheme, separable ordinal groupings (here varying along Dimension 1) should have some independence from the identities of the items themselves (here varying along Dimension 2). B) Poor order sensitivity, conversely, results in a lack of separability on the basis of position. C) Ordinal positional coding is not the only explanation for separation of items on the basis of order. Repetition suppression effects, for example, may also produce separation. However, in this case the suppression of response magnitude creates steadily smaller confidence ellipses when moving from item 1 to 2 to 3, likewise with steadily smaller distances between their centroids.

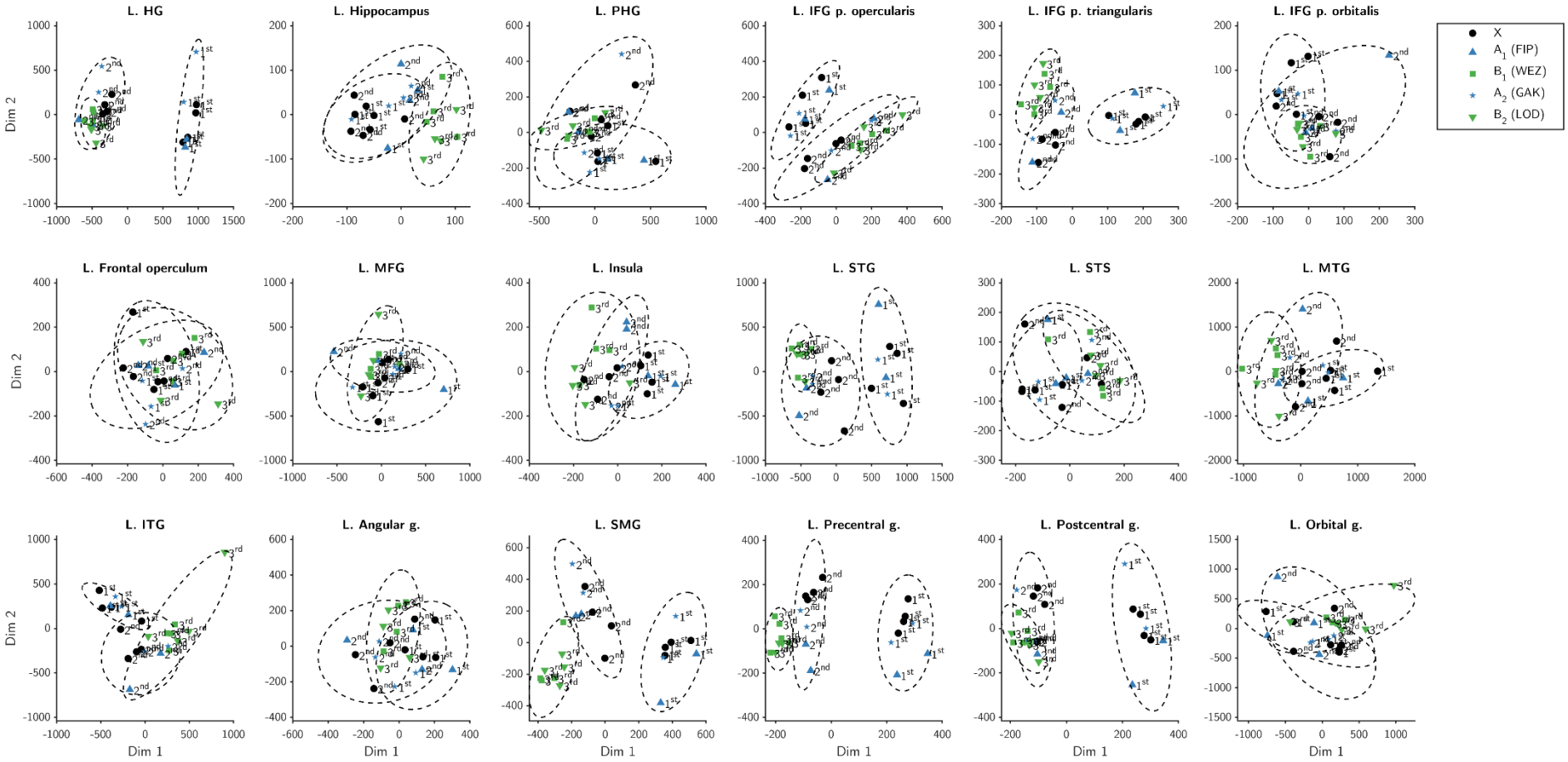


Figure 5.6: **MDS on Euclidean distances between syllabic responses across time.** Syllables lasted 600 ms, followed by an ISI of 150 ms. ERPs across trials were computed for each syllable from data under each of the 8 factorial conditions of the experiment, windowed at 0-750 ms around syllable presentation, producing a 24 × 24 RDM for each region. Points are coloured according to syllabic identity and labels include the ordinal position of the syllable in the sequence. CI_{95} ellipses are drawn around each set of syllables of the same ordinal position (all the first, or second, or third items of a sequence). The ellipses separate well into 3 distinct groups in regions that likely incorporate ordinality into their responses. In this way, when consulting the regional plots, evidence compatible with ordinal encoding can be seen in IFG pars triangularis, SMG and precentral gyrus.

Figure 5.6 shows the empirical two-dimensional MDS results derived from Euclidean distance matrices, comparing responses to syllables in all positions of the presented sequences. Each point represents the response to a syllable from a sequence presented under one of the experimental conditions. Results are shown for every region of interest in the left hemisphere. Confidence ellipses depict the 95% confidence boundary around points of each ordinal category (syllables occurring in the 1st, 2nd and 3rd sequence positions; each point's label includes its ordinal position as a numeral).

Consulting the MDS results and the schematics in Figure 5.5, it can be seen that responses to the syllables separated in a configuration potentially compatible with ordinal coding in left IFG pars triangularis, supramarginal gyrus (SMG), and precentral gyrus. In particular, the responses in left precentral gyrus suggested relatively detectable independence between the separable ordinal positions (varying along the horizontal axis) and item identity (varying along the vertical axis), strongly aligning with the predicted "ordinal coding" schematic. By contrast, in the right hemisphere, only superior temporal gyrus (STG) exhibited clean separation (see Appendix 2: Supplementary figures, Figure 0.11). A number of regions appeared to be more similar to the predicted repetition suppression response, for example left Heschl's gyrus and STG, which exhibited some separation, but steadily decreasing confidence ellipse areas and inter-centroid distances over time. The majority of remaining regions in both hemispheres appeared simply to exhibit poor order sensitivity, however.

Some clustering could also be observed on the basis of item identity. Hippocampus and precentral gyrus arguably showed the clearest separation between non-salient (**X**; black circles) and salient (**A**/**B**) syllables (however, no region showed clean separation on the basis of salience; CI_{95} ellipses not shown). No regions showed clean separated of **B** syllables on the basis of grammaticality (CI_{95} ellipses not shown).

5.4.3 Dynamic representational dissimilarity

Figure 5.7 shows a whole brain Procrustes-aligned multidimensional scaling (pMDS) plot for a single participant ("423"; see Chapter 3), produced by conducting the pMDS analysis on all contacts of the participant, rather than region by region. The plot depicts multidimensional scaling (MDS) results over time, where MDS output at each time point was derived from the condition-condition RDM for the 8 experimental conditions. This was the only passively recorded participant of the cohort, and so these results had the potential to reveal activity that was uncontaminated by task-

related motoric demands. Consequently, data in this plot was windowed up to 4000 ms post-stimulus instead of the 3000 ms used in later analyses. Annotations clarify the representational geometry in low-dimensional space. Moving chronologically, several key features can be observed.

Firstly, it can be seen that patterns during sequence presentation (0-2100 ms) clustered in low-dimensional space according to the identity of the newly presented syllable. This immediately demonstrates that the multivariate ECoG patterns contained enough information to recover the identities of the presented items using the pMDS approach. The traces also revealed reorganisation of the representational geometry at times not directly aligned with the presentation of an auditory stimulus, for example coincident with the inter-stimulus interval separating syllables 2 and 3 of the stimulus sequence (1350 – 1500 ms, where representational signals diverged on the basis of whether a sequence began "Ax..." or "xA..."); and beyond the end of sequence presentation.

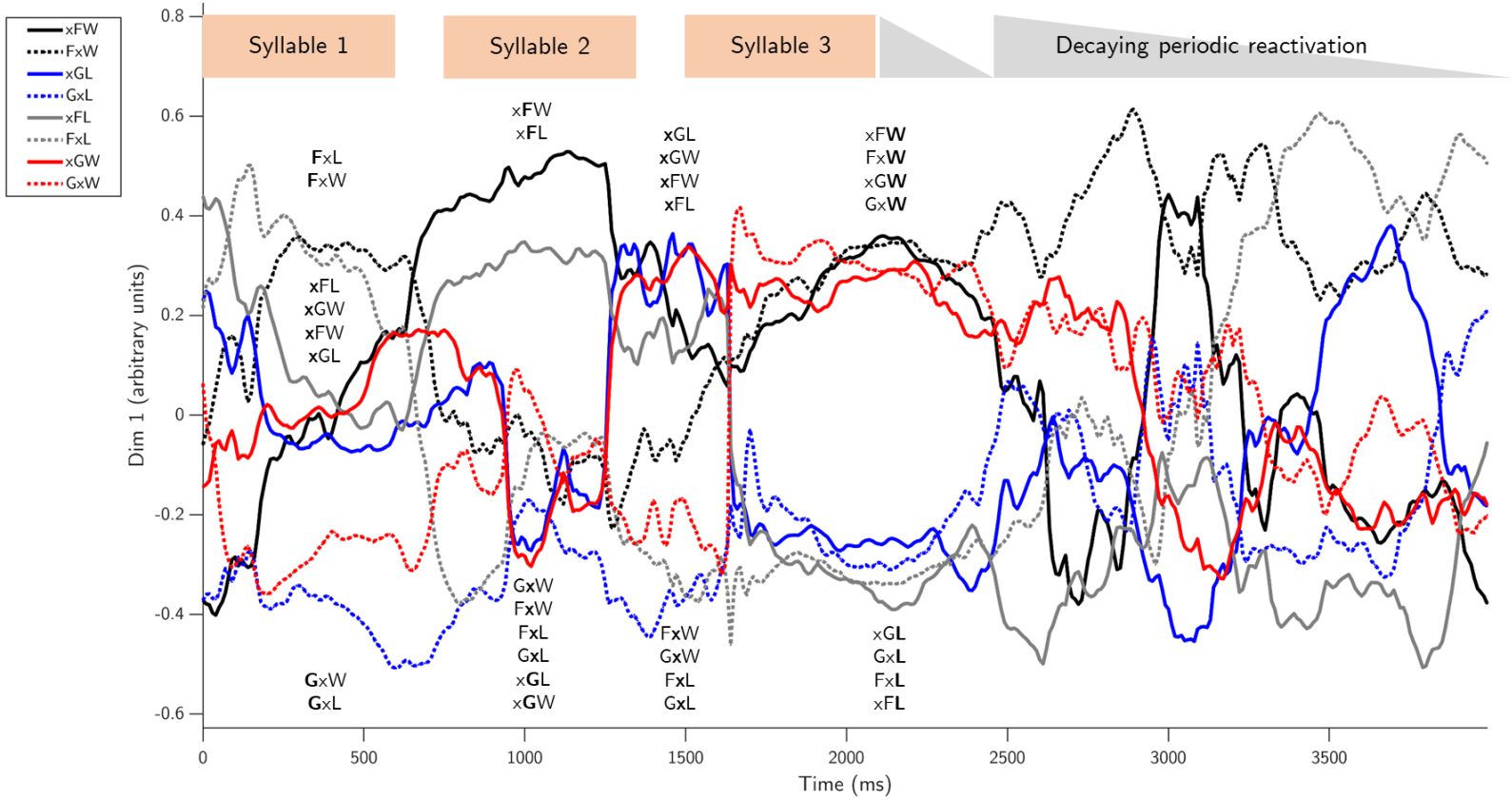


Figure 5.7: **pMDS results using "whole-brain" aggregation of recordings in an exemplar participant.** The participant (423) ran a passive task configuration. Stimuli were presented at the times shown in the orange boxes overhead. The ordinate axis depicts representational space. Lines separate along this dimension when multivariate responses diverge, and cluster when similar. Annotations on the plot correspond to nearby lines, to clarify the clustering of patterns over time. Here "x" refers to the class of **X** items, F to "Fip" ($\mathbf{A_1}$), G to "Gak" ($\mathbf{A_2}$), W to "Wez" ($\mathbf{B_1}$) and L to "Lod" ($\mathbf{B_2}$). Representational clustering can be observed relating to the identity of each newly presented syllable. Between second and third syllables, "**Ax**..." sequences are clearly distinguished from "**xA**..." sequences. Following presentation, a long period (~400ms) of decay is observed for the third syllable, after which further repeated reorganisation of the representational geometry can be observed at a frequency of approximately 3 Hz (delta). Each oscillatory period, meaningful separation of the sequences deteriorates, potentially suggesting a process of decaying, oscillator-mediated reactivation.

Post-sequence, the prominent **B** response decayed to undetectable levels, but the representational traces continued to fluctuate, exhibiting a periodic reorganisation at around 3 Hz until at least the end of the analysed window. Because this whole-brain visualisation is dominated by the strongest signals across the brain at any time, the decay of **B** does not conclusively indicate that the syllable no longer elicited a response in selected regions and contacts. However, the post-sequence periodic reorganisation indicates that exogenous or endogenous oscillatory processes predominated. Representations separated at the maxima of these periodic fluctuations were seemingly not random, but instead appeared to be related to each other over time. To avoid additional labels, this portion of the plot (2500 – 4000 ms) does not include annotations, but consulting the key will convince the reader that the clustering pattern at each periodic maximum is related to that of the previous cycle. Specifically, at the first post-stimulus peak, patterns still largely separated on the basis of the identity of the **B** element, but with each subsequent period, separation became increasingly disordered, losing all discernable relation to the task by around 4000ms. This suggests that the post-stimulus activity reflects a decaying delta-mediated reactivation of the **B** response, or else a steadily decaying **B** response overlaid with some other strongly delta-mediated process.

Overall, the pMDS plot succinctly depicts whole-brain multivariate dynamics. Similarly, regional pMDS results (not shown) depicted representations resolved across space and time. However, it is not especially useful to summarise information in this way at the cohort level; and it is secondly difficult to track condition-condition contrasts. As a result, the pMDS traces were simplified to dynamic representational dissimilarity (DRD) plots as described in 5.3.8 (and especially Figure 5.2). The cohort mean DRD **A**, **B** and "grammaticality" traces for every ROI are plotted in Figure 5.8. The results for left Heschl's gyrus are enlarged in Figure 5.9. Note that only data between -150 and 3000 ms were included in this analysis, and in subsequent causal analyses. Signals after 3000 ms were potentially informative, but omitted to avoid confounding the putative grammaticality signal with motor activity during the response period; despite between-participant counterbalancing of response button handedness, differences in motor responses would likely covary with sequence grammaticality in appropriately responding participants.

Figure 5.8: **Dynamic representational dissimilarity (DRD) results for each ROI across the cohort (n = 12).** Bilateral Heschl's gyri instantiate strongest responses to **A** in both the first and second sequence positions (0 to ~1500ms), and **B** in the third position (~1500 to 3000ms). Other ROIs showing stimulus-driven response peaks include left: hippocampus, frontal operculum, and STG; and right: IFG pars opercularis and triangularis, insula, and SMG.

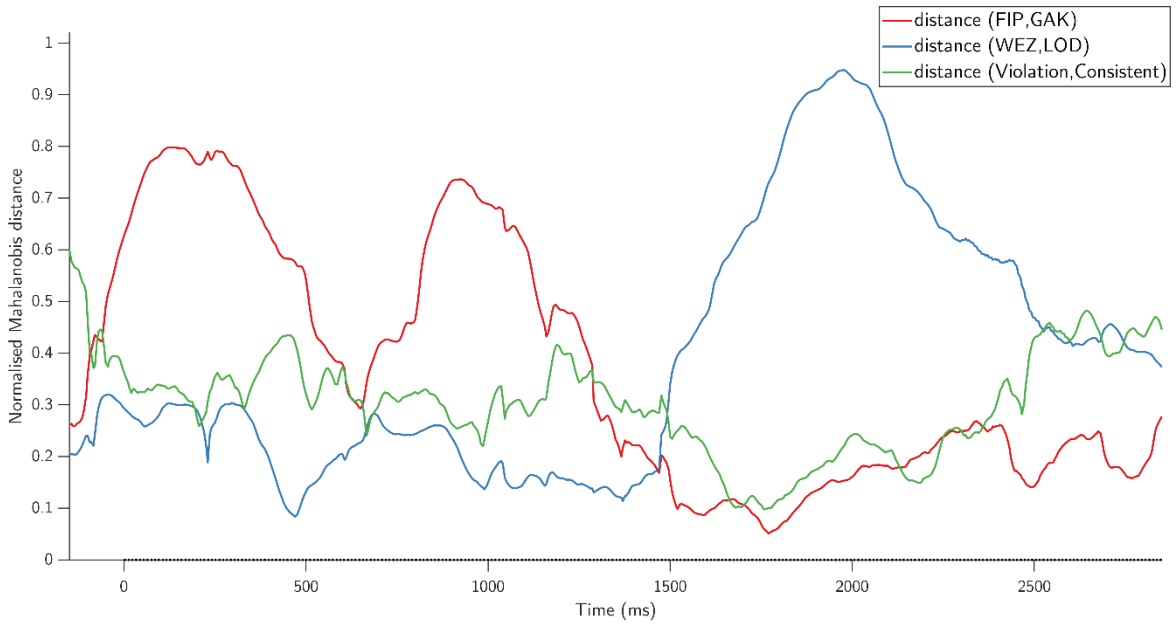


Figure 5.9: **An enlargement of the left Heschl's gyrus DRD results (A, B, grammaticality).** The signal at each time point reflects a 300 ms window. Data is plotted aligned with the centre of each window. The analysis omits data after 3000 ms to exclude signals during the motor response period. Responsiveness to **A** elements (red) peaked twice (0 to ~650 ms, peaking at ~200 ms, and 650 to ~1700 ms, peaking at ~900 ms). Each peak corresponds to **A** in the first position or **A** in the second position of the sequence, as dictated by the grammar of the task. The response to the **B** element exhibits a later peak (blue; ~1500 to 3000 ms, peaking at 1950 ms). Sequence playback ends at 2100 ms. The group mean "grammaticality" signal (green) fluctuates far less interpretably, but peaks towards the end of each syllable, relative to the preceding few hundred milliseconds (or, conversely, is suppressed slightly during syllabic presentation, most clearly during presentation of the third syllable; difference not significance-tested).

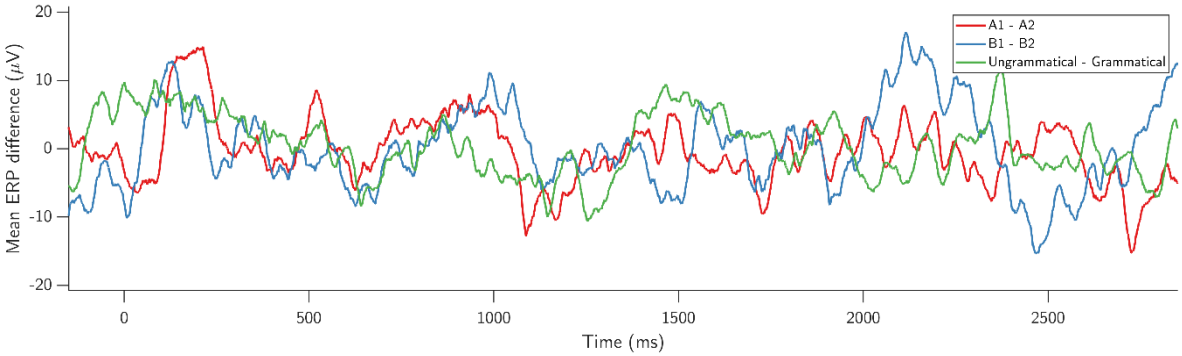


Figure 5.10: **Difference waves for grand mean (cohort-level) event-related potentials (ERPs) under opposing conditions in left Heschl's gyrus.** Note that these results rely on time-domain information, as in the previous figure. However, this more traditional analysis cannot make use of the high-dimensional information available across multiple sites and time points. As a result, even though this ERP analysis and the DRD analysis both act on the same dataset, it can be seen that item sensitivity can only be discerned through the multivariate DRD approach, whilst the encodings remain obscured following the mass univariate approach.

As in the pMDS visualisation, the DRD plots revealed clear periods of sensitivity to presented items (see Figure 5.8). In particular, bilateral Heschl's gyri exhibited large peaks in the dissimilarity waves at the approximate times of syllable presentation. The **A** item DRD wave exhibited a double-peaked contour, where each peak aligned with the presentation of the first and second syllables. This can be attributed to the presence, across all trials, of **A** items in both the first and second positions of the presented sequence (as in **A**x**B** or x**AB**; see 2.3.3). The **B** visualisation, however, featured a single peak due to its consistent presentation time (always 1500-2100 ms). Finally, the "grammaticality" signal appeared to locally peak towards the end of each syllable, relative to its level at the time of maximum syllabic encoding, most noticeably after auditory playback ended. Note that this observation is based purely on the group mean shown, and no statistical significance should be surmised. Even so, the possibility that this might be a signature of a long latency process originating beyond auditory cortex motivated the subsequent Granger causality analysis (subsection 5.3.9).

At this point, it is important to emphasise that the time-domain DRD approach is more powerful than simply subtracting two event-related potentials (ERPs). This is because the data at each time point reflects content across a spatiotemporal history of neural responses, not just a response at a single point in time and space. To highlight that the DRD provided information that would not be apparent in a basic ERP study, equivalent cohort-level ERP contrasts were produced, reproduced in Figure 5.10. It can be seen that the multivariate results (Figure 5.9) revealed task-related effects that were not revealed with any fidelity under the univariate ERP contrast.

Returning again to the DRD plots in Figure 5.8, it can be seen that many regions other than Heschl's gyrus appeared to instantiate encodings that covaried with the presented syllables. Focussing on the **A** and **B** DRD waves (red and blue respectively), we identified a number of regions outside of Heschl's gyrus that appeared to exhibit stimulus-driven activity at times roughly correspondent with the presentation of two or more of the syllables. These included left hippocampus, left superior temporal gyrus (STG), left frontal operculum, right supramarginal gyrus (SMG), right inferior frontal gyrus (IFG) pars opercularis, right IFG pars triangularis, and bilateral precentral gyri. Many regions also included less transparently stimulus-driven activity, which still likely constituted task-related effects. For example, in the left hemisphere, the hippocampus, IFG pars triangularis, frontal

operculum, superior temporal sulcus (STS) and angular gyrus all featured what appeared to be recapitulations of the **A** and **B** responses at ~2500-3000 ms, well after sequence playback ended.

Finally, although the previous DRD figures depicted **A**, **B** and "grammaticality" contrasts, also computed were "1st syllable" ($\mathbf{S_1}$) and "2nd syllable" ($\mathbf{S_2}$) contrasts. A "3rd syllable" ($\mathbf{S_3}$) contrast is exactly identical to **B**, because the third syllable was always a **B** item. Figure 5.11 shows cohort-level $\mathbf{S_1}$, $\mathbf{S_2}$, $\mathbf{S_3}$ and "grammaticality" contrasts in left Heschl's gyrus. Crucially, each syllable exhibited a single, distinct peak. This enabled the flow of information to be more accurately tracked during subsequent causal analysis, since it can be seen that the $\mathbf{S_1}$ and $\mathbf{S_2}$ syllabic responses (which might easily exhibit different causal flow patterns) were no longer amalgamated as they were in the double-peaked **A** contrast.

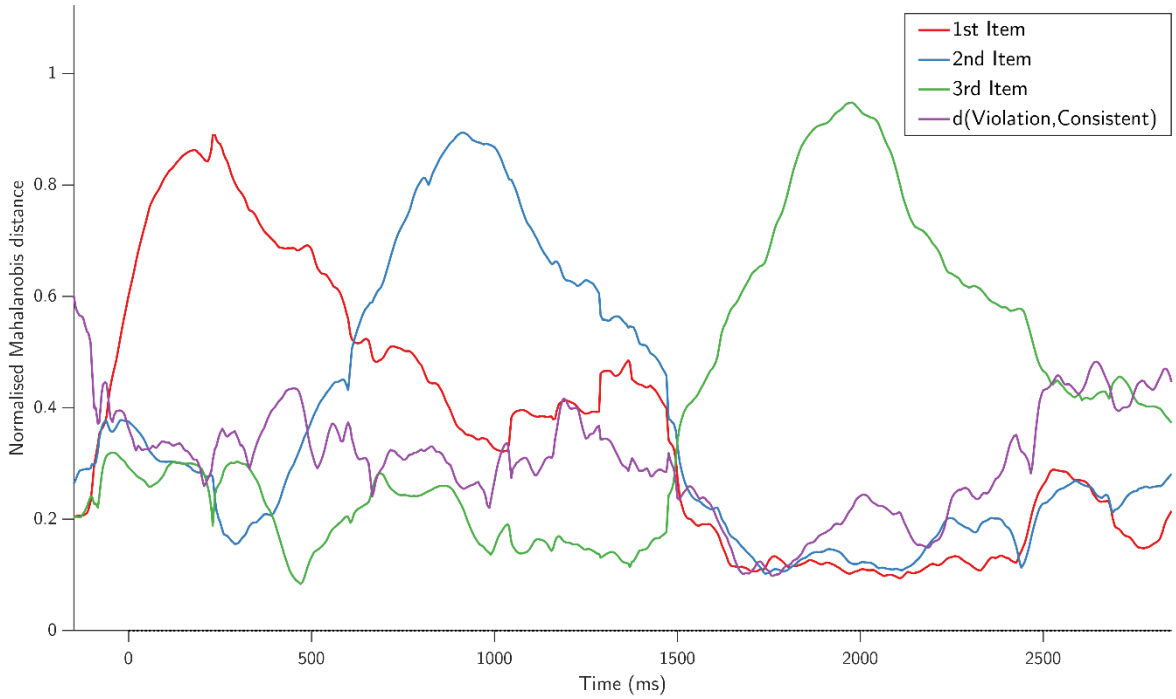


Figure 5.11: **An enlargement of the left Heschl's gyrus DRD results (1st, 2nd, 3rd syllables, grammaticality).** This concords with the **A** and **B** contrasts, but highlights that, by taking the maximum distance between 3 categories ($\mathbf{X/A_1/A_2}$), DRD waves could be plotted for each syllable separately.

5.4.4 Representational Granger causality

Because of our specific interest in the mechanistic functioning of the language network, and the left language dominance demonstrated by the cohort during pre-implantation assessment (see Chapter 3), representational Granger causal (GC) analysis was restricted to left-hemispheric regions.

The analysis again yielded five sets of causal results: $\mathbf{S_1}$; $\mathbf{S_2}$; $\mathbf{S_3}$ (i.e. **B**); **A**; and "grammaticality". These summarised, across the cohort, the causal relationships between regions with respect to each of the time-resolved DRD contrasts. The raw results of these five analyses (Granger causality strength; *p*-values; and significance for each pairwise comparison between regions) are presented in Figure 5.12. It can immediately be seen in all five cases that significance was met for a subset of region-region associations (bottom row, showing significance masking, with black cells denoting significance). Moreover, by visual inspection, it may be appreciated that the significance mask in each case was asymmetrical, in that relatively few of the causal associations were significant in both directions simultaneously. This was confirmed quantitatively by computing a "symmetry" metric (Lanzi, 2016) on the mask, defined as:

$$s = \frac{\left|\frac{1}{2}(A + A^T)\right| - \left|\frac{1}{2}(A - A^T)\right|}{\left|\frac{1}{2}(A + A^T)\right| + \left|\frac{1}{2}(A - A^T)\right|}$$

where A denotes the matrix, A^T its transposition, and $|\cdot|$ the matrix norm. The matrix symmetry metric is a scalar value in the range -1 to 1, where -1 indicates "complete anti-symmetry" and 1 indicates "complete symmetry".

For every causal analysis, the symmetry was close to zero ($\mathbf{S_1}$: 0.094; $\mathbf{S_2}$: 0.17; $\mathbf{S_3}$/**B** item: 0.19; **A** item: 0.015; grammaticality: 0.16). This indicated no especially strong symmetry or anti-symmetry; that is, it indicated a sparse pattern of largely unidirectional causal flow between regions. By comparison, a traditional Granger causality analysis conducted on regional grand mean event-related potentials across all trials, downsampled to the same equivalent sampling rate of 40 Hz, revealed many more significant causal links than the representational GC runs (see Appendix 2: Supplementary figures, Figure 0.12), and in a significantly more bidirectional pattern (as indicated by a symmetry of 0.53; one-sample *t*-tested against the representational GC symmetries, yielding $p < .001$). Although GC on all trials is already a rather different concept to GC on difference waves, these results begin to indicate that, as previously suggested, the representational GC tracks the flow of relatively targeted types of information compared to traditional causal analysis and that, as expected, such flow is relatively directed compared to causal analysis on raw neural responses.

Figures 5.13, 5.14 and 5.15 show causal graphs computed in 5.3.9 from the representational GC results. Within each plot, nodes represent regions, and the edges between any two regions are proportional in width to the magnitude of the net significance-masked causal flow between them.

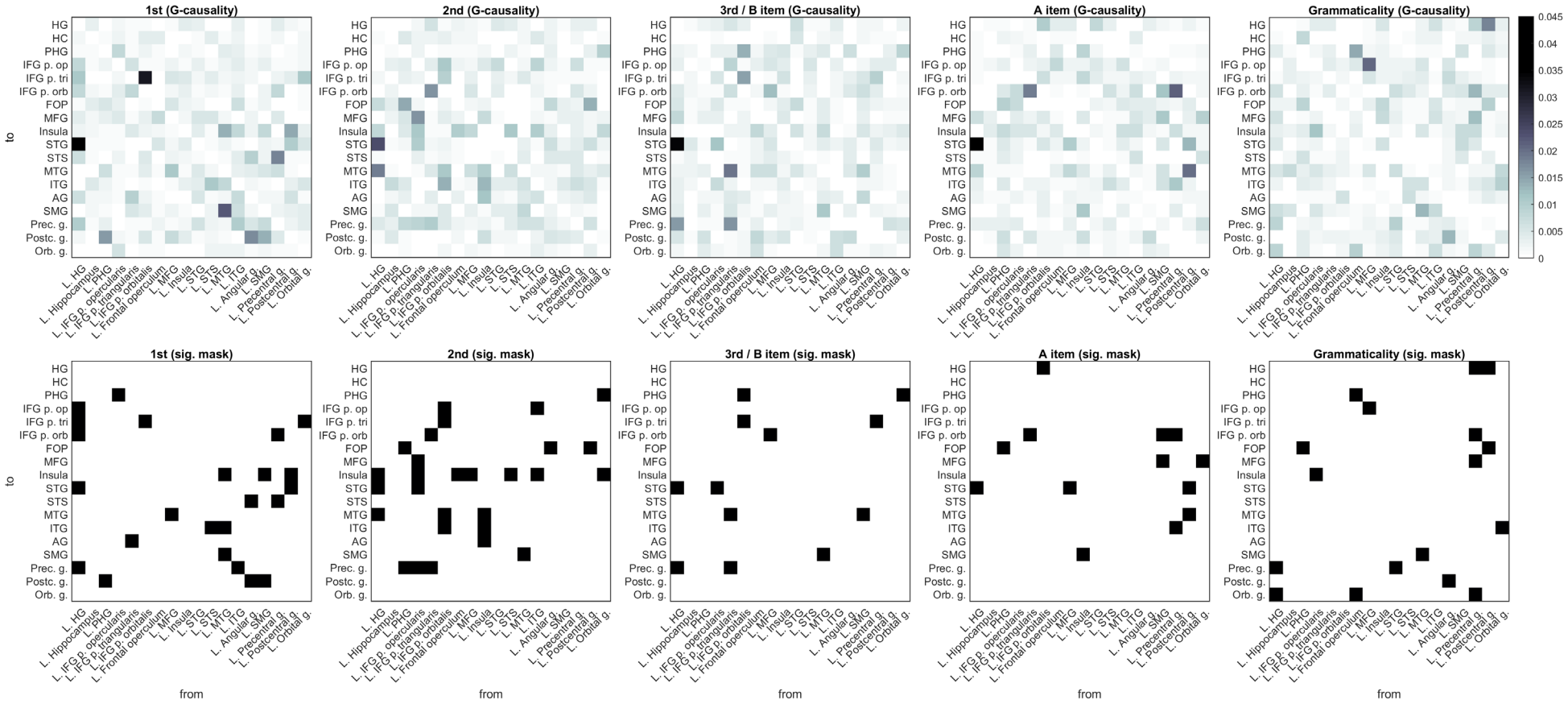


Figure 5.12: **Raw representational GC results (G-causality strength, top, and significance masks, bottom).** The horizontal axis shows causal origins, and the vertical axis shows destinations. Masks show significance at $p = .05$ (FDR-corrected for multiple comparisons; black denotes significance). From left to right: syllabic $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$ (**B** item) flow; **A** item flow; "grammaticality" flow. Many significant links are unidirectional, as can be seen from the fact that the significance masks show a non-symmetrical pattern across the downward diagonal. Note that causal associations differ between representational datasets. Syllabic, **A** and **B** data shows flow from Heschl's gyrus to STG, for example, whilst the grammaticality signal shows significant flow from STG to precentral gyrus. "Grammaticality" features the greatest number of significant outbound links from frontal operculum.

Additionally, the diameter and colour of each node depicts the *total significant causal flux* of a region. The total significant causal flux was defined as the total significance-masked flow out of a region minus the total significance-masked flow into a region, across all regions. This can help clarify the overall flow of information through the network. Here, red nodes are net sources (those where total significant causal flux is positive), and blue nodes are net sinks (i.e., recipients) of causal flow (those where total significant causal flux is negative). We refer to this value simply as causal flux from this point onwards. Heschl's gyrus is circled for emphasis. The layered layout of each graph was automatically optimised to reveal hierarchical structure using MATLAB's "layout" function (which implements Gansner et al., 1993; Barth et al., 2004; Brandes & Köpf, 2002).

Figure 5.13 shows the causal flow graph produced from the "first syllable" ($\mathbf{S_1}$; top panel) and "second syllable" ($\mathbf{S_2}$; bottom panel) DRD waves. Figure 5.14 includes the corresponding causal flow graph for the "third syllable" DRD wave ($\mathbf{S_3}$ or **B**; top panel). As expected of signals that covaried with the encoding of the auditory stimuli, the largest positive causal flux in all three cases emanated from primary auditory cortex, and the single strongest causal link for every syllable was from Heschl's gyrus to superior temporal gyrus (STG). This finding provides initial validation for the approach. However, there was evidence of some isolated nodes that acted as net sources of syllabic content, but with no inward links from auditory cortex. It is possible that these regions were not passing syllabic content, but rather other information that covaried with it. However, a more likely explanation is that the direct or indirect causal links to these regions from auditory areas simply did not meet the significance threshold. This is borne out by the raw GC results shown in Figure 5.12 (top row).

There were other substantial commonalities between the syllabic causality networks. All three causal graphs involved significant causal flow to and from all subregions of IFG, in particular an interaction between IFG pars triangularis and IFG pars orbitalis that was present in all three syllables' causal flow networks, although it varied in strength and causal direction in each case. Both second and third syllable responses appeared to flow directly from IFG pars opercularis to STG, and whilst a direct association was absent in the case of the first syllable, there was still indirect net flow between these two regions in the same direction. In the case of all three syllables, significant flow was detected running from middle temporal gyrus (MTG) to supramarginal gyrus.

The lower panel of Figure 5.14 shows the causal flow for **A** items overall. This signal was previously shown to peak twice in some areas, including primary auditory cortex, with peaks corresponding to the encoding of the first and second syllables (see 5.4.3). Despite this fact, causal analysis of the **A** signal could be justified by the fact that this was the only DRD wave likely to covary chiefly with the identity of the first grammatically salient item in the artificial grammar sequence. That is, the **A** signal contrasted only **A** items, but the syllabic signals also incorporated data on the extraneous **X** items.

The **A** item causality network has immediate similarities to the individual syllabic networks in the form of causal outflow from Heschl's gyrus to STG, and interactions between IFG pars triangularis and orbitalis. However, here, whilst Heschl's was still a net causal source, it also accepted significant causal inflow from IFG, specifically IFG pars orbitalis and, indirectly, IFG pars triangularis, SMG, and precentral gyrus. Consulting the raw G-causality matrices once again (Figure 5.12, top row), it can be seen that there was some similar, but weaker causal association from IFG to Heschl's for the second and third syllabic signals, but this fell below threshold significance, and in those cases was not chiefly from IFG pars orbitalis as seen here (Figure 5.14, lower panel). Furthermore, the raw GC results under the syllabic contrasts show little evidence for the SMG/precentral gyrus to IFG flow seen here. Overall, therefore, there are notable sources of inflow to Heschl's from IFG, and to IFG from SMG, that are markedly different from those seen in the individual syllabic graphs.

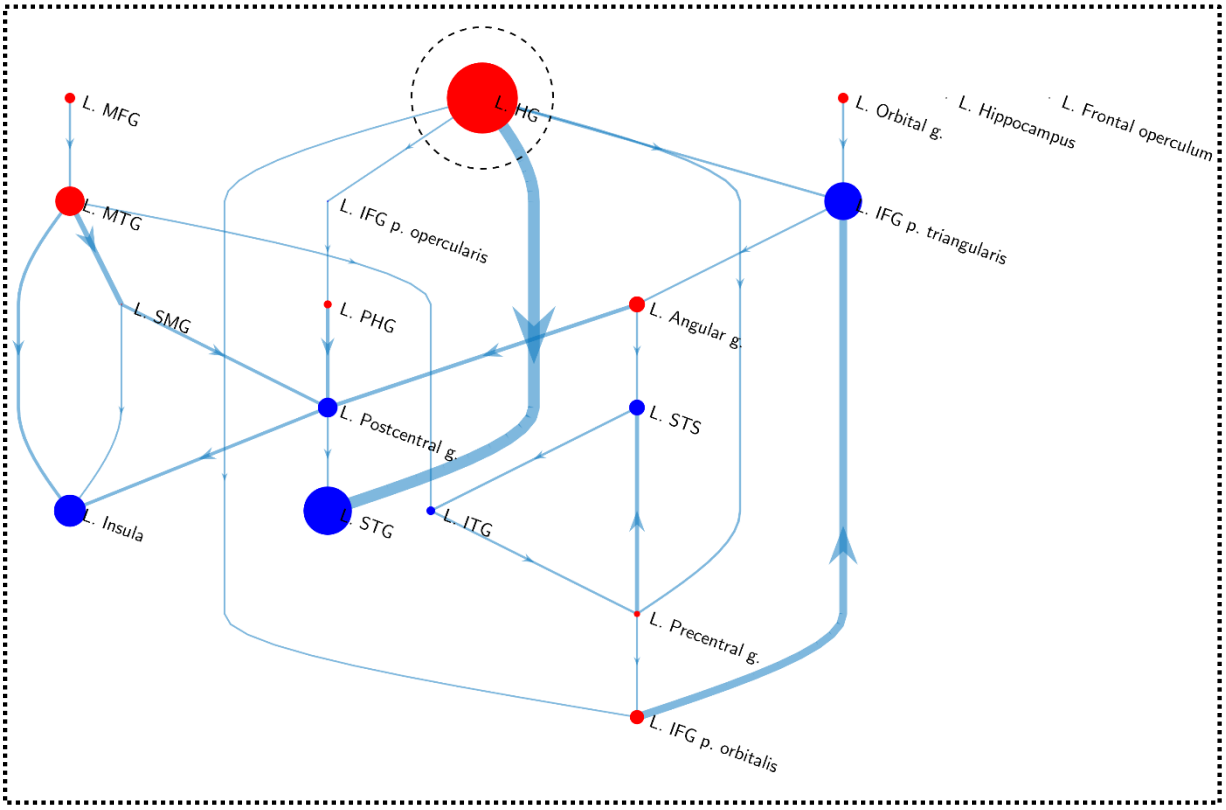
Whilst significant causal outflow from Heschl's is expected, *inflow* of **A** to Heschl's gyrus seems an unlikely finding. Therefore, although discussed further in the next section, it is timely to consider briefly how this could happen. It is possible that the **A** signal, by amalgamating information in two sequence positions (x**A**... and **A**x...), yields spurious results in causal analysis. This cannot be ruled out, due to the complex nature of the interactions in each dataset. It is also possible that, by eliminating the responses to **X** items, the **A** signal reveals otherwise obscured flow patterns that fell below statistical significance in the first and second syllabic graphs, although the raw results do not provide especially strong evidence for this. Most intriguing, however, is the possibility that the **A** graph is contaminated by a covarying signal: specifically, a prediction of **B**. Assuming the participant understands the task and attempts to match **B** to **A**, the presentation of **A** should elicit some

prediction of the correct **B**. Because the correct **B** is entirely dependent on the presented **A**, these two signals would, under the most ideal circumstances, precisely covary across all trials. This would not be the case for information in the first or second syllable DRD waves. It is therefore possible that the **A** item graph is contaminated with "predicted **B**" signals.

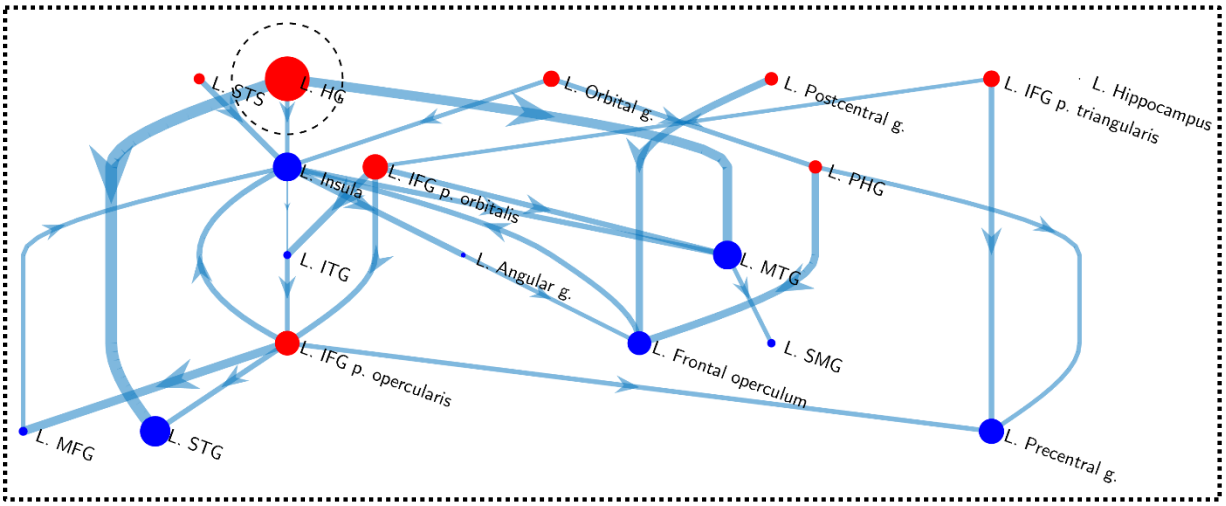
Finally, Figure 5.15 shows the causal flow of the "grammaticality" DRD wave. As with all of the DRD waves, in the absence of further testing, the statistical significance of the distances at each time point cannot be inferred. This fact is especially relevant to the "grammaticality" signal, where fluctuations were previously revealed to be extremely modest in comparison to the syllabic DRD waves. As a consequence, the Granger causality results on this signal should be interpreted with particular caution, as the physiological nature of the underlying signal cannot be guaranteed. However, it is also important to bear in mind that any significant causal flow under this analysis is statistically inferred at the group level, and thus represents replicable flow patterns across the participants. Here, Heschl's gyrus was a net recipient of causal flow (shown as a blue node), receiving significant direct inflow from precentral and postcentral gyri, and indirect flow from angular gyrus and STG. Areas of ventrolateral prefrontal cortex including frontal operculum, IFG pars opercularis and pars orbitalis were all significantly involved in the causal network, whereas IFG pars triangularis, interestingly, was not significantly associated. Frontal operculum here showed bidirectional causal associations with parahippocampal gyrus (PHG) and significant causal output to orbital gyrus. Somewhat unexpectedly, none of the significant causal associations ran from ventrolateral prefrontal cortex to lower-order regions including primary auditory cortex. However, the raw GC results (Figure 5.12, top row) do show very weak direct and indirect causal outflow from IFG pars opercularis and triangularis to Heschl's gyrus, which did not meet significance.

In all of the representational causal networks, including the **A** and "grammaticality" graphs, there existed interactions between areas of ventrolateral prefrontal cortex and PHG. However, no statistically significant causal links were detected to or from the hippocampus itself in any of the causality networks. Note that, as for all the representational GC results, this implies a lack of causal flow for the putatively specific representations we assessed using this novel method (A/B, syllables and grammaticality), not a lack of causal flow in general, which was revealed using the traditional GC analysis on raw LFPs mentioned previously (Figure 0.12, Appendix 2: Supplementary figures). In the GC analysis on raw LFPs, significant causal inflow was detected to the hippocampus from STS,

MTG, and STG, whilst significant causal outflow was detected from the hippocampus to STS, MTG, PHG and IFG pars opercularis. Overall, this suggests a possible role for the hippocampus under the AGL task, but seemingly using representations that are distinct from those that we traced through the frontotemporal network. Moreover, the fact that the LFP-GC analysis revealed outflow from the hippocampus to the PHG, whilst the representational GC analysis revealed significant causal associations with the PHG alone, suggests that the PHG potentially acted as an intermediary between a distinct hippocampal system and the wider frontotemporal network under this task.

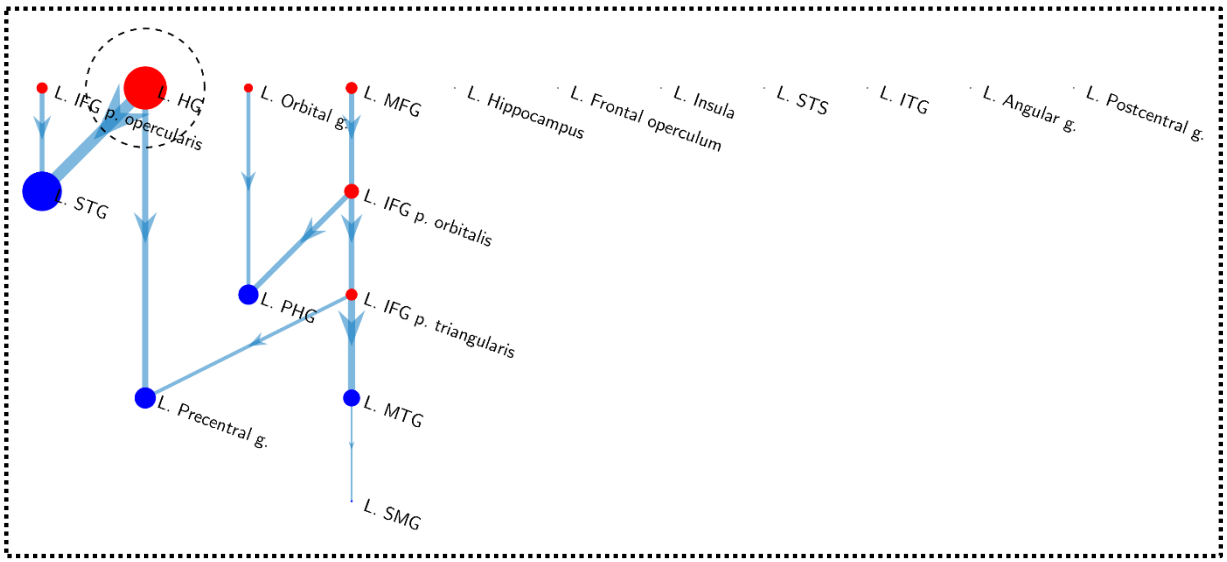


▲ First syllable (S_1)

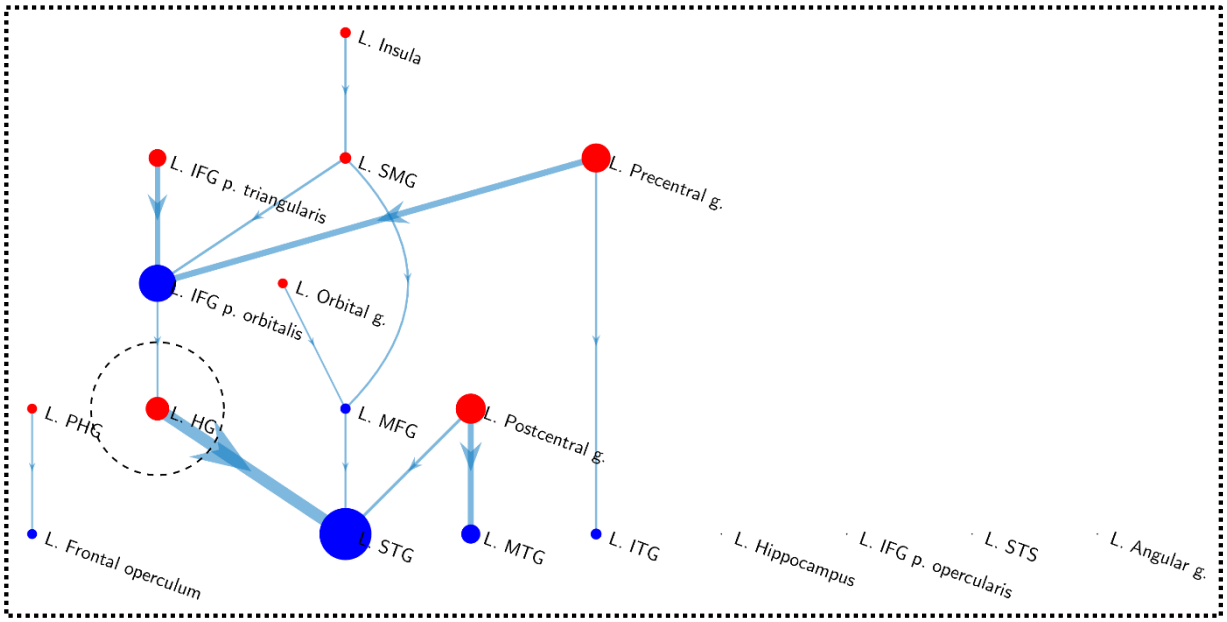


▲ Second syllable (S_2)

Figure 5.13: **Representational flow graph, first and second syllables.** Here, nodes depict regions, edge weights depict net causal flow between them, and node colour and diameter indicate the direction and size of the total causal flux to/from a region (red: overall source; blue: overall sink). Heschl's gyrus is a prominent overall source of the first and second syllabic content, as we would expect, exhibiting strong causal outflow to superior temporal gyrus (STG). Note that multiple causal generators (nodes in red, with outflow only) are possible because the graph only shows net flow, and because some causal associations between sub-networks might not reach significance, isolating them.

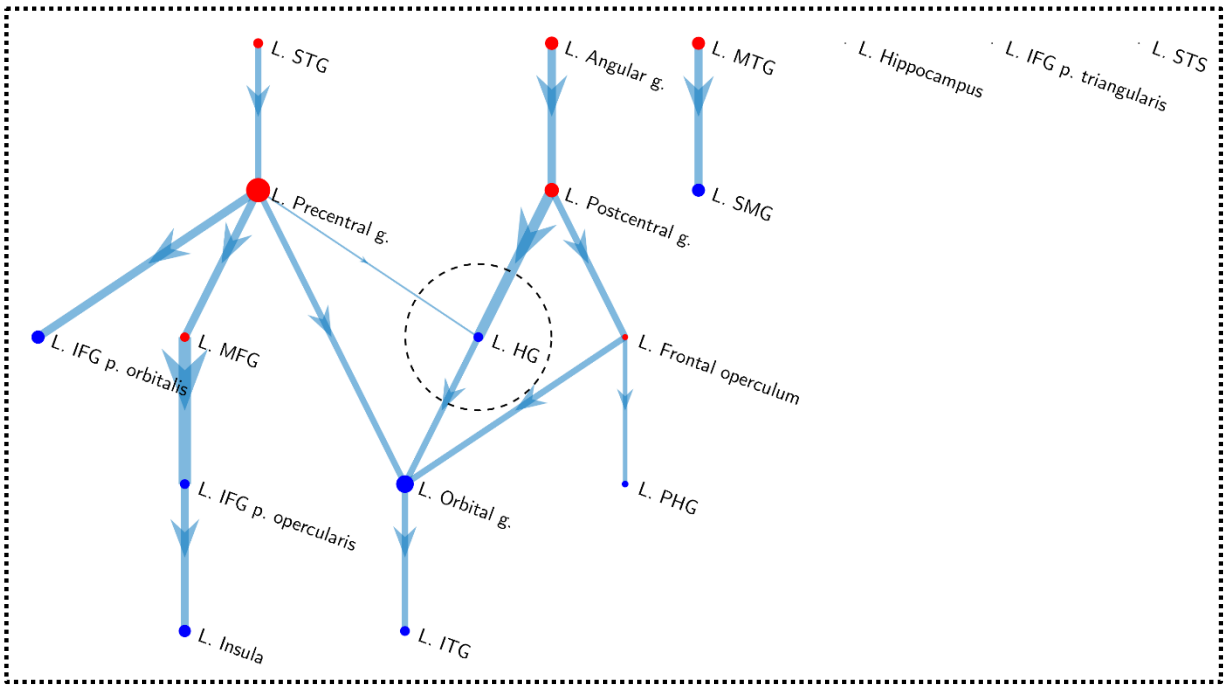


▲ Third syllable (S_3; "B" item)



▲ "A" item

Figure 5.14: **Representational flow graph, A and B items.** Again, nodes depict regions, edge weights depict net causal flow between them, and node colour/diameter indicate direction/size of total causal flux to/from a region (red: source; blue: sink). **Third syllable (B item), top:** Heschl's gyrus is again an overall source with significant outward causal flow, notably to superior temporal gyrus (STG). **A item, bottom:** As in previous figures, Heschl's is an apparent source of **A** signals, with significant outward causal flow to superior temporal gyrus (STG). However, there is also inflow to Heschl's from IFG, SMG and precentral gyrus, unlike the the $1^{st}/2^{nd}/3^{rd}$ syllable flow patterns. Since any accurately *predicted* **B** would covary with the *presented* **A**, the flow to Heschl's may represent feedback of a predictive **B** representation, confounding the **A** flow pattern, but concordant with the predictive coding hypothesis.



▲ "Grammaticality"

Figure 5.15: **Representational flow graph, "grammaticality".** More than any other representational Granger causality result reported here, note that the configuration of this causal flow network should be interpreted with caution due to the low signal-to-noise ratio of the underlying DRD wave. However, here, the flow pattern is markedly different from that of the other graphs. Notably, Heschl's is now an overall sink (blue), rather than a source, and appears to receive prominent causal flow from precentral and postcentral gyri, with STG and angular gyrus as the earliest detected sources. Whereas parahippocampal gyrus was detected as an overall causal source for **A** items, as well as first and second syllable content studied separately, it acts as a causal sink here, as was also the case within the **B** item flow graph. This is also the only flow graph in which frontal operculum was detected as being significantly causally associated with other regions whilst also being an overall source, rather than a sink.

5.5 Discussion

We conducted a battery of multivariate analyses on the AxB dataset and revealed evidence for regional encodings consistent with our computational model, as well as novel findings on the time-resolved instantiation and inter-regional flow of representations during sequence processing. The outcome of each analysis is discussed further below.

5.5.1 Neural pattern similarity recapitulates existing functional connectivity findings

The neural pattern similarity (NPS) analysis revealed undirected functional connectivity consistent with that shown under a previous functional connectivity study (Wang et al., 2020), motivating further analysis of multivariate data collected under our task. In particular, the NPS analysis replicated existing findings on the functional heterogeneity of subregions of inferior frontal gyrus (IFG; Clos et al., 2013), motivating the continued study of representations in anatomically distinct subregions of IFG. This heterogeneity has been shown previously to relate to distinct patterns of cytoarchitectonic and receptoarchitectonic microstructure (Zilles & Amunts, 2018).

Our NPS results also provided evidence for communication (in the form of relatively high pattern similarity) between IFG and the temporoparietal junction, an interaction critical for language in humans and likely involved in the integration of both syntactic and semantic information (Friederici, 2011, 2012). Thus, under our AGL task, we observed meaningful patterns of functional connectivity between regions comprising a broad fronto-temporal/parietal language network.

5.5.2 Comparison of syllables across time reveals ordinal coding in selected regions

In our comparison of neural responses to syllables across times and conditions, we found evidence potentially compatible with ordinal coding in three left-hemispheric regions: IFG pars triangularis, SMG and precentral gyrus; and in one right-hemispheric region (STG). Overall, this accords with the recently reported findings of Henin et al. (2021), who determined that ordinal coding is prominent in word-responsive sites across the brain. However, their task involved segmentation of a continuous syllabic speech stream, whilst in our task, every word was a single syllable separated by pauses of 150 ms duration, in sequences surrounded by seconds of silence. Because our sequence items were all monosyllabic, we cannot contrast syllabic and word position coding under this task, though this is a promising avenue for future research. Nonetheless, the fact that we too found evidence compatible with an ordinal coding account might suggest that the

ordinal code is not exclusively relevant to rapidly unfolding syllables within individual words, but extends to longer timescales, potentially encapsulating the order of individual words in a sentence. This accords with eye-tracking results in infants and adults suggesting that both continuous speech and speech delineated by pauses elicit ordinal encoding, though this tendency appears to be enhanced in the presence of pauses (Fló, 2021).

A number of prior findings have highlighted the general importance of ordinal codes in humans and nonhuman primates (reviewed in Dehaene et al., 2015), which motivated their prominent inclusion within our computational model of sequence processing, VS-BIND (Calmus et al., 2019). However, we added to the body of existing evidence by revealing regional encodings that were not inconsistent with site-specific predictions of our model. Signs of potential ordinal coding in precentral gyrus were consistent with VS-BIND, which suggests domain-general roles for motor and premotor cortex in linear sequencing. This assertion was based on previously reported findings including studies by Carpenter et al. (2018) and Ninokura et al. (2004), who both revealed ordinal sensitivity within these regions in the brains of rhesus macaques during serial order recall tasks. Under our task, one subregion of ventrolateral prefrontal cortex, IFG pars triangularis, also exhibited signs of ordinal coding, in line with the predictions of VS-BIND. Within our model, representations arising in IFG are theorised to incorporate ordinal coding in a manner that supports both linear and hierarchical structure-building. We could not provide any evidence of hierarchical structure-building in neural responses to this task. However, this would not have been unexpected even in the highest performing participants, because sequences were only three items long, and the dependency of interest (**A→B**) was a simple relationship over a variable distance.

Based on our model, which posits that non-salient items are filtered out before the remaining items are structured, we did expect some segregation of the **A**/**B** and **X** representations depending on their variability, chiefly in frontal operculum (FOP). We previously asserted that, since FOP appears to be sensitive to item and/or chunk frequency (Rong et al., 2018; Friederici, Fiebach, et al., 2006), it potentially therefore has a role in salience filtering (Calmus et al., 2019). We did not see evidence of salience filtering in this region under this task, however, although there was some limited evidence of this segregation in the hippocampus and selected other regions (as shown in 5.4.2). In particular, the ordinal representation detected in precentral gyrus appeared to segregate the **X**s from other items. In general, however, there was little evidence that uninformative **X** items were filtered out at an early

stage and completely discarded from later structures as posited in VS-BIND. This could suggest that the model requires revision. However, an alternative possibility is suggested by the behavioural data (previously reported in Chapter 3), which showed that, at the cohort level, the implanted participants were unable to learn the rules of the artificial grammar above chance levels. It is possible that, in this cohort, the reason performance was so poor was because non-salient **X** items were erroneously considered salient by the participants, resulting in their incorporation into underlying encodings, disrupting attendance to the truly salient **A** and **B** items, and confounding learning of the sequencing rules. Henin et al. (2021) obtained similarly poor behavioural confirmation, theorising that it might relate to hippocampal deficits specific to the temporal lobe epilepsy population. However, here, this explanation lacks parsimony, because poor performance might have resulted from many environmental or iatrogenic factors. Indeed, we previously presented evidence that, in a healthy cohort, even self-reported fatigue correlated with reduced performance on this task (Chapter 2).

A final point on this analysis regards the nature of the ordinal code. A possibility not explored by Henin et al. (2021) is that the "ordinal coding" detected by this type of analysis might not represent an explicit ordinal code (as envisaged in our model) at all, but rather any effect that covaries with position. This includes simple repetition suppression (where repeated exposure to a stimulus attenuates the neural response, and which we explored in Figure 5.5) and expectation suppression (where expectations of stimulus occurrence attenuate the observed response). However, we consider this to be an unlikely explanation for our key multidimensional scaling (MDS) results in left SMG, IFG pars triangularis and precentral gyrus, for two reasons. Firstly, repetition and expectation suppression, both observed in primary auditory cortex (Todorovic & Lange, 2012), likely explain aspects of the responses we observed in Heschl's gyrus (Figure 5.6, p. 197), yet representations in this region were not observed to segregate effectively on the basis of position. Secondly (and also visible in the Heschl's gyrus results) is the fact that, as previously stated in 5.4.2, under the influence of any form of suppression, multivariate representations move steadily closer together in Euclidean representational space as they diminish in magnitude. In primary auditory cortex, this resulted in a large confidence ellipse around the first items, but markedly shrinking ellipses as one progressed through the sequence, presumably as a result of expectation suppression (see Figure 5.6). Repetition suppression based on the frequency of the items is less easily excluded, but, in this case, rare **X** items would be expected to elicit responses of a greater magnitude than the common **A** and **B** items, which

would likely result in increased dispersal of the **X** items in the MDS visualisation, relative to **A** and **B**. These were not features of the representations in IFG pars triangularis, SMG or precentral gyrus. Overall, this suggests that explicit ordinal coding is a plausible explanation for the results we observed. Further studies, varying the length of sequences, could reveal support for our assertion that these codes describe boundary-relative, rather than absolute, ordinal position (see Chapter 4).

5.5.3 Dynamic representational dissimilarity reveals peri- and post-stimulus activation

The Procrustes-aligned multidimensional scaling (pMDS) and dynamic representational dissimilarity (DRD) analyses, whilst conducted on time-domain data, revealed sensitivity to the stimuli that could not be discerned from a simple univariate ERP contrast. Specifically, the DRD analysis clearly revealed stimulus-driven activity in a number of regions across the left-hemispheric frontotemporal language network, including auditory cortex, the hippocampus, left frontal operculum and precentral gyrus. This last result was concordant with the earlier finding that an ordinal sequence representation was instantiated in precentral gyrus (see above).

As well as auditory encodings, the DRD analysis sought to isolate "grammaticality" signals, expectation violation signals relating to the proto-syntactic structure of the stimulus sequences. The group mean grammaticality DRD in Heschl's gyrus appeared to peak following each syllable in comparison to the preceding few hundred milliseconds (see Figure 5.9), although this effect was not statistically tested. More importantly, the group-level dissimilarity waves were not statistically analysed to determine whether they exceeded any particular baseline, making it possible that some of the fluctuations in a low amplitude dissimilarity signal fell below the natural noise floor of the data. However, due to the relative latency of the putative peaks in the group mean "grammaticality" wave, and the nature of the underlying contrast (responses to violation versus consistent sequences), it was considered that it could represent a signal originating outside auditory cortex, potentially describing "prediction error". Trial-by-trial encoding of prediction errors has previously been observed in auditory cortex during presentation of auditory oddball sequences (Rubin et al., 2016). However, although a predictive coding account of this data was seemingly supported by the subsequent representational Granger causality analysis (see below), two additional caveats must be borne in mind when interpreting the "grammaticality" signal.

Firstly, some regions, including FOP, exhibited prominent early peaks in the grammaticality waves that could not have been elicited by true grammatical sensitivity. However, FOP in particular had limited coverage across the cohort (n = 2 participants had coverage of left frontal operculum by 3 contacts in total, and n = 1 participants had coverage of right FOP through a single contact). It is therefore possible that the DRD waves in the FOP had a relatively low signal-to-noise ratio. Here, it is evident that it would be beneficial in future to refine methods to statistically test the DRD waves against a silent or "noise only" baseline, although given the intended sensitivity of the multivariate contrasts to latent dynamics, defining this baseline period may not necessarily be trivial.

Secondly, it is known that only one of the 12 participants achieved significantly higher than chance performance on both runs of this task. Although the normalised DRD waves showed grammaticality peaks at various task-related times at the cohort level, it is possible that the chance-performing participants did not exhibit any particular sensitivity to the sequences, and that a weak signal in those individuals was driven by artefactual components. However, there are reasons to believe that, across the cohort, the grammaticality wave at least partly comprised a physiological signal of syntactic relevance after the time of the grammaticality manipulation. Firstly, group-level electrophysiological analyses collectively revealed time- and time-frequency domain effects in left IFG pars opercularis, right MTG and right hippocampus (Chapter 3). Furthermore, in the subsequent causality analysis (see next section), statistically significant causal outflow of the grammaticality response was detected at the cohort level emanating from regions including angular gyrus, again concordant with a role for the temporoparietal junction in the detection of syntactic violations. Irrespective of the DRD results, however, the AxB task was undeniably a cognitively demanding task, as evidenced by our behavioural findings in the neurosurgical cohort (Chapter 3), and also foreshadowed by bimodal performance on the task in healthy participants (previously reported in Chapter 2). Future work will include the application of multivariate methods to data collected under a less challenging and/or more meaningful AGL paradigm, and with a greater emphasis on methods supporting the statistical validation of the DRD wave.

Overall, the pMDS/DRD approach produced output of a high temporal resolution, seemingly capable of tracking the dynamics of relatively specific cognitive representations on the millisecond scale. This can be achieved on ECoG data using decoding approaches (for example, to decode inner speech; Martin et al., 2018), which learn to recognise patterns in the data by using classifiers that

require training prior to analysis. The DRD approach, however, is an *en*coding analysis (see Kriegeskorte & Douglas, 2019 for a comprehensive comparison of encoding and decoding methods). Unlike a decoding analysis, no training process was required, making the approach computationally efficient. The algorithm also has few tuneable parameters, simplifying its use. One critical tuneable parameter, however, is the length of the temporal window used to construct the spatiotemporal vectors over time (see 5.3.8). In this analysis, we used a window of 300 ms, but integrating neural responses over longer or shorter time periods might reveal different dynamic patterns. The determination of an "optimal" timescale is relevant to many analyses including decoding analyses. However, it is likely that the question of which timescale to integrate over is ill-posed, since information at different timescales encodes different features (for example, in the auditory system, Teng et al., 2016), because neuronal timescales appear to be heterogeneous even within single regions (Cavanagh et al., 2020) and because they are to some extent functionally dynamic and therefore ever-changing (Gao et al., 2020). In future, however, further work might be undertaken to empirically determine the window of integration providing the most information about a particular signal. Indeed, formal methods have proposed for just this purpose, for example by analysing the entropy of motifs in fMRI across specific windows to determine the most appropriate decoding timescale (Deco et al., 2019). In this case, Deco et al. (2019) determined that the richest dynamical repertoires across the brain were observed on a timescale of around 200 ms (as opposed to scales of different orders of magnitude), extremely close to the temporal window we used.

5.5.4 Representational Granger causality reveals distinct flow patterns for different types of dissimilarity wave

The representational Granger causality analysis appeared to be a promising tool for the analysis of neural data, revealing sparser causality networks that were significantly more unidirectional than Granger causal (GC) analysis across all trials of equivalently downsampled ECoG data. This suggested the waves tracked relatively well-isolated representations that exhibited characteristic flow patterns across the cohort during sequence processing. However, the specificity of the DRD waves was not guaranteed, and any signal covarying with the signal of interest would also have been revealed. The presence of covarying signals was especially relevant to the interpretation of the **A** item causal flow, which we return to shortly.

Individual syllabic representations were revealed to flow through a broad network of regions encompassing critical parts of the language network. In all cases, this included STG, all subregions of IFG, and parts of the temporoparietal junction, consistent with existing findings on the role of the language network in non-linguistic structured sequence processing, and known dorsal and ventral connectivity between STG and IFG (Petersson et al., 2012; Fitch & Friederici, 2012). Causal associations with frontal operculum (FOP), however, were not revealed except in the case of the second syllable. Since the original DRD wave shows a prominent encoding of each item in FOP, this most likely represents a failure to reach statistical significance due to the lack of coverage in this area.

Similarly, representational GC on **A** and **B** DRD waves revealed causal flow largely consistent with the individual syllabic findings, which all revealed prominent outflow from primary auditory cortex as expected. However, although Heschl's gyrus was detected as a net causal generator of the overall **A** item signal, analysis also revealed significant inflow to primary auditory cortex from precentral gyrus, IFG pars triangularis and SMG, seemingly by way of IFG pars orbitalis. This potentially represented unanticipated tracking of a covarying "predicted **B**" representation (see 5.4.4). This result accords with previous findings indicating that increased in IFG pars triangularis and orbitalis reflect the instantiation of syntactic predictions (Matchin et al., 2017). However, a carefully designed paradigm would be required in order to definitively distinguish covarying **A** and "predicted **B**" representations.

Finally, we reported causal flow of a supposed "grammaticality" or expectation violation signal, computed from the multivariate differences between neural responses to consistent versus violation sequences. It was considered that cohort-level causal analysis of this inconsistent grammaticality signal could be meaningful because, despite individual variation, we would expect causal flow to be coherent across the cohort if the signal tracked a mechanistically specific process. However, although we suspect this to be the case, without also statistically verifying that individual DRD waves breached their respective noise floors, it is still possible for the group-level Granger causality analysis to have revealed effects even if it were conducted on noise. As well as statistically testing each DRD wave against some baseline, an approach that could mitigate this possibility in future would be to generate the DRD waves using a noise-resistant method of dimensionality reduction in place of simple multidimensional scaling, for example "MDS+" (Peterfreund & Gavish, 2021).

Acknowledging the aforementioned limitations, causal analysis of this signal revealed that auditory cortex was a net recipient of causal flow at the group level, which was detected as ultimately emanating from auditory association cortex and the temporoparietal junction. Frontal operculum, believed to be involved in local dependency violation (see Wilson et al., 2017; Friederici, 2011), was significantly causally associated with other regions here, acting as an overall source, but generating significant causal outflow to orbital gyrus and parahippocampal gyrus (PHG), as well as also receiving flow from PHG and precentral gyrus. However, the extreme limitations of frontal opercular coverage in this dataset must be acknowledged when interpreting its roles and causal associations. Interestingly, this network also included causal inflow to IFG pars opercularis (a region also incorporated in the causal networks for $\mathbf{S_1}$ to $\mathbf{S_3}$), which we previously found to respond differently to grammatical and ungrammatical adjacent sequences at the group level (see Chapter 3).

We consider that the top-down causal flow exhibited by the putative "grammaticality" signal, in conjunction with the **A** item causal findings, is not incompatible with a hierarchical predictive coding account (Summerfield & de Lange, 2014; Friston & Kiebel, 2009; Rao & Ballard, 1999). In this account, an individual region of a hierarchical network continuously produces context-driven predictions of the representations in lower order regions, which are passed down to them, yielding a prediction error response. Prediction errors, in turn, propagate to higher-order regions. These same prediction-error interactions occur at each level of the hierarchy. Whilst individual errors pass from lower-order (sensory) regions to higher-order (integrative) regions, our evidence very tentatively suggests that predictions, as well as specific prediction-error interactions, may both appear to propagate downwards as we track a specific expectation violation signal (for further investigation of time-resolved perspectives on predictive coding, see Hogendoorn & Burkitt, 2019). However, further statistical and/or empirical evidence is required to definitively demonstrate that the "grammaticality" signal is not driven by noise, but rather reflects task-driven processing amenable to causal analysis.

5.6 Conclusion

In this study, we undertook a battery of multivariate analyses on ECoG data collected during an artificial grammar learning task, with the aim of clarifying regionally instantiated neural representations and inter-regional information flow during sequence processing operations. Neural

pattern similarity revealed inter-regional response similarities that mirrored previously reported patterns of functional connectivity derived from fMRI (Wang et al., 2020). Analysis of syllabic representations across time revealed evidence for ordinal coding in IFG pars triangularis, SMG and precentral gyrus, compatible with our model (Calmus et al., 2019) and complementing and clarifying recent findings by Henin et al. (2021). Dynamic representational dissimilarity, derived from multivariate patterns over time, revealed sensitivity to sequential stimuli in a host of regions including areas of ventrolateral prefrontal cortex and the hippocampus, which manifested as peri-stimulus and post-stimulus activation of sequence representations. Causal analysis of these signals revealed distinct flow patterns between regions of the language network for different facets of the artificial grammar learning (AGL) stimuli, suggesting prominent involvement of IFG in both organised stimulus encoding and syntactic prediction, as well as a key role for areas of the temporoparietal junction in the propagation of syntactic predictions and possible syntactic violation signals. Collectively, these analyses begin to provide a glimpse of the complex interplay between components of the language network, and argue strongly for a productive interaction between model- and data-driven approaches.

Chapter 6. General Discussion

Neuroimaging and behavioural results have implicated a frontotemporal-parietal network of brain regions in both language and domain-general structured sequence processing operations (Friederici, 2011). However, the representations instantiated by these regions require clarification, as do the functional interactions between them, if we are to better understand their role in cognition (Dehaene et al., 2015; Petkov & Wilson, 2012). Moreover, we require clarity not just on the mechanisms of human sequence learning, but also on the limits of human capabilities, which have yet to be firmly established (ten Cate et al., 2020).

The studies reported in this thesis aimed to clarify the behavioural and neuronal sensitivity of subjects to violations of expected sequence structure, and to clarify the contributions of regions implicated in the construction of cognitive sequence representations. In particular, I aimed to clarify the representations and dynamic interactions arising in and between key regions of the frontotemporal language network. This included inferior frontal gyrus (IFG, incorporating the site commonly known as *Broca's area*; Broca, 1861), which still has a much-debated and relatively underspecified role considering its long-appreciated involvement in language processing (Yan et al., 2021; Matchin & Hickok, 2020; Hsu et al., 2017; Molenberghs et al., 2012; Pulvermüller & Fadiga, 2010; Rogalsky et al., 2008; Hagoort, 2005; Friederici, 2002; Grodzinsky, 2000; Smith & Jonides, 1999; Caplan & Waters, 1999; Rizzolatti & Arbib, 1998).

In service of these aims, I undertook analysis of behavioural (Chapter 2; Chapter 3) and electrophysiological data (Chapter 3; Chapter 5) collected under an auditory artificial grammar learning (AGL) paradigm, the *AxB* task, analysing the results using univariate techniques and novel multivariate approaches, as well as computationally modelling neural representations relevant to sequence processing (Chapter 4). Ultimately, I sought to triangulate existing and novel findings in order to characterise site-specific representations instantiated during sequence processing.

Through behavioural analysis of data collected on the AxB task in healthy adults, I have shown that, at the group level, participants were able to concurrently learn flexible adjacent/non-adjacent dependencies in the presence of statistical cues to salience, and that even in those who failed to exhibit high performance on a grammaticality judgement measure, reaction times revealed sensitivity to the grammar. Conversely, by analysing behavioural collected in a cohort of neurosurgical patients

undertaking an extremely similar version of this AGL task, I have demonstrated that summary grammaticality judgement performance and reaction time measures did not provide evidence for learned sensitivity to the grammar in the multifactorially affected cohort. However, analysis of intracranial data in this cohort revealed an event-related potential (ERP) effect under this task concordant with a key canonical syntactic component, the LAN, in inferior frontal gyrus (IFG) pars opercularis, and time-frequency effects in the right middle temporal gyrus and hippocampus. Despite this, univariate group-level analysis of ERP and time-frequency results provided limited insight into site-specific mechanistic roles. On this basis, I have suggested that multivariate – rather than univariate – analyses have the potential to reveal greater insights into the mechanistic foundations of sequence processing, and have subsequently demonstrated this by presenting novel methods for multivariate analysis and applying them to the intracranial *AxB* dataset. This has, in particular, suggested a role for the predictive coding framework in a coherent account of frontotemporal sequence processing.

From a neurocomputational perspective, I have proposed that sequence processing hypotheses be constrained by integrating existing evidence into large-scale, mechanistic models. Specifically, I have motivated an approach to modelling the brain using concepts from the field of hyperdimensional computing. Using this approach, I have defined a model of key aspects of sequence processing, VS-BIND (Calmus et al., 2019), that proposes specific neurocomputational mechanisms and generates a number of predictions ripe for testing. Finally, as a result of the aforementioned multivariate analyses on the *AxB* neural data, I have found evidence to support a number of predictions of site-specific codes specified within VS-BIND, in particular providing evidence that both IFG pars triangularis and precentral gyrus instantiate sequence representations integrating ordinal positional codes.

At a broad level, the goal of this schedule of work was to improve our understanding of the human brain's internal perspective on temporal structure. I have contended that this appears to shape our perception and production of actions, music and language, and potentially cognition in general. However, although one of the motivating arguments for this contention is the shared importance of hierarchy between domains, I have not directly addressed the encoding of hierarchical relationships in my experimental work, rather forming neurocomputational hypotheses about hierarchy solely within VS-BIND and focussing experimental efforts on an understanding of simple adjacent and

non-adjacent structures. The nature of hierarchical encodings is therefore a principal challenge yet to be fully addressed within our schedule of work. However, there are a number of other aims and hypotheses that arise naturally from the findings reported in this thesis. Here, I discuss the specific challenges addressed by my work, describe how my results contribute to the understanding of sequence processing in the frontotemporal network, and propose additional research that could clarify the neural codes employed in these and other regions, thereby advancing our understanding of the mechanistic foundations of sequence processing in the human brain.

6.1 Strengths, limitations and future avenues

6.1.1 Human performance in a mixed dependency AGL task

Findings under the mixed dependency AGL task in healthy adults (Chapter 2) revealed that participants could learn the explicit relationships in the task, but, given that they formed a bimodal distribution, also revealed that it was less trivial to achieve a high score on the explicit grammaticality judgement task than one might have assumed, even with the assistance of perceptual cues. To some extent, however, the suggestion that the task was non-trivial aligned with a number of prior reported studies incorporating non-adjacent dependencies (see 2.2.1). Moreover, even if the healthy participants were not able to perform well on the grammaticality judgement task, they appeared to show sensitivity to the sequences in the form of reaction time effects. Nevertheless, given the likely multifactorial effects on the patient cohort of their clinical condition, prescribed interventions and medication, it is perhaps the case that, in hindsight, performance was insufficiently high in the healthy cohort to allow for any reasonable chance of observing sustained high grammaticality judgement performance in the patients. Moreover, reaction time effects are likely to be severely undermined or eliminated by the prescription of opioid analgesics and other powerful drugs used to control pain in the patient population (Rogers et al., 2013; Spear et al., 1992). As a consequence, future studies would benefit from the use of alternative online measures of sequence learning, especially passive measures. Outside of traditional electrophysiological contrasts, one increasingly popular analysis that has been suggested to track sequence learning is neural frequency tagging (NFT; Batterink & Paller, 2017), which quantifies neural entrainment effects at the frequency of words relative to the frequency of individual syllables. Learning of structured content is associated with enhanced word-level entrainment relative to responses to random content. This approach has formed the basis for other compelling analyses of sequence processing (Henin et al., 2021; Ding et

al., 2016), but we were unable to retrospectively apply it to our data because of the jitter we had deliberately introduced to our trials to minimise the impact of motor activity on measured responses (see 3.3).

6.1.2 Electrophysiological signatures of sequence processing in the human brain

Despite a lack of behavioural confirmation, findings under the electrophysiological *AxB* task in neurosurgical patients (Chapter 3) revealed a significant event-related potential (ERP) effect concordant with a canonical finding under linguistic syntax-violation tasks: specifically, a group-level difference between responses to adjacent grammatical and adjacent ungrammatical sequences in left IFG pars opercularis, strongly resembling the left anterior negativity (LAN). Time-frequency analyses were less intelligible, revealing relatively unexpected gamma-band effects in the right middle temporal gyrus (MTG) and hippocampus – although, as we noted, both of these regions, or else their left hemispheric homologues, have previously been implicated in processes pertinent to artificial grammar learning (Opitz & Friederici, 2003; Loui et al., 2011). Unfortunately, the cohort size was insufficient to allow statistical inferences to be made at the group level in a number of regions (see 3.3). This means that it is possible that right MTG and right hippocampus were not the most grammatically-responsive regions, despite the results of the group analysis. Similarly, this restriction makes it very difficult to surmise hemispheric differences, since the number of regions reaching mimimum acceptable levels of coverage in the left hemisphere exceeded those in the right. Finally, for the same reason, it is not possible to judge the relative importance of the contributions of different subregions of ventrolateral prefrontal cortex in relation to IFG pars opercularis, since IFG pars triangularis, IFG pars orbitalis and frontal operculum both had insufficient coverage to reach α. This restriction was caused by the limited coverage of specific regions, and ultimately this limitation was a function of the availability of clinical patients. In future, a focus on simplifying the cognitive demands of the task and broadening the acceptable regions of interest might increase the rate at which suitable participants can be recruited, since we excluded a small number from further study due to pre-implantation performance and coverage concerns (see 3.4.1).

A second point of interest is the implantation procedure itself. Given that the highest performing implanted participant was in fact a stereoelectroencephalography (sEEG) patient, rather than a recipient of a subdural ECoG array (see Figure 3.1), it is possible that a less invasive method of recording is a pre-requisite for high performance on this kind of task. It has previously been shown,

for example, that patients undergoing sEEG experience significantly less pain and require fewer opiates compared to those implanted with subdural grids (Y.-C. Wang et al., 2020). Depending on their availability, sEEG patients may therefore prove to be a more appropriate population in which to undertake more cognitively demanding tasks such as the *AxB* task in future.

Finally, whilst we randomised the trial order under the *AxB* task, we did not undertake any other balancing to control trial ordering. Thus, it was possible to obtain sequences of several trials where the grammaticality of the sequence remained identical to that of the previous trial, and this did happen on one or two occasions, and a run of >7 identical conditions was observed on more than one occasion. Given that refamiliarisation is typically incorporated into AGL tasks precisely because learned dependencies may be corrupted by repeated exposure to incorrect sequences (Kuppuraj et al., 2018), this may have had the effect of introducing a spurious sense of "violation" at unanticipated times when a sequence under a different condition was introduced, akin to effects observed under the *local-global task*, a variant of the oddball task (see Chapter 1) where repeated presentation of a deviant sequence can result in any new transition to a "grammatical" *target* sequence instead generating a mismatch response (Marti et al., 2014). However, as an argument against this effect, preliminary analyses (not reported here), could not ascertain any clear associations between the condition(s) of the prior 1-2 trials and effects in the subsequent trial.

6.1.3 Neurocomputational hypotheses on combinatorial binding and sequence processing

VS-BIND proposed specific roles for a number of cortical regions, and the hippocampus, in sequence processing. In particular, we treated the hippocampus as a flexible sequence playback engine (see Chapter 4). At present, neither direct nor indirect hippocampal-IFG interactions are defined in our model, but the time- and time-frequency results we revealed under an AGL paradigm (Chapter 3 and above), which revealed significant effects in both regions, highlight the fact that it is likely to be assistive to define and test falsifiable neuro-computational hypotheses on such interactions in future.

Excitingly, multivariate analyses of left IFG, precentral gyrus and supramarginal gyrus (SMG) suggested that they may all instantiate ordinally tagged sequences of items (Chapter 5). IFG and precentral gyrus, specifically, were predicted by us to contain such representations on the basis of an integrative view of existing research (Chapter 4). SMG did not specifically enter into our predictions,

primarily because VS-BIND was not intended to incorporate aspects of semantics, and temporoparietal areas have putative roles in lexical-semantic integration (Thompson et al., 2015). It may therefore be premature to seek to incorporate SMG into VS-BIND given its stated purpose as a non-linguistic account of sequence processing.

A goal that must certainly be prioritised as a result of the neurobiological data is the comprehensive integration of predictive coding aspects (Friston & Kiebel, 2009; Rao & Ballard, 1999) into VS-BIND. Presently, the model encompasses mechanistic manipulations of cognitive structures that have putative roles in specific sequence processing operations, and lacks a formalisation of predictive coding aspects, although this problem was alluded to in Chapter 4.

There are two ways in which predictive coding might be implemented in VS-BIND. One is to conceive of the predictive coding mechanism as a sub-symbolic process that is completely distinct from the symbolic representations themselves. Under this approach, the idealised representations would continue to be conceived of as at present, but predictive coding would be introduced through the sub-symbolic behaviour of individual artificial neurons, for example by implementing the full model using populations of virtual neurons within a large-scale simulation framework such as "Nengo" (Bekolay et al., 2014). Parts of VS-BIND have already been implemented within Nengo (see Figure 0.10, Appendix 2: Supplementary figures), and some existing research exists attesting to the feasibility of implementing predictive coding under this simulation framework (Ororbia, 2019; though note that their implementation was not neurobiologically constrained).

The second way in which predictive coding could be implemented is directly at the level of the symbolic representations. This is somewhat intuitive, since, under this approach, regions of the model would explicitly instantiate "prediction" structures and "error" structures. These would have the same canonical forms but represent different states of the system. One reason that this form of implementation is potentially attractive is that it provides an "oven-ready" description of exactly what representations regions such as IFG would be manipulating. To be specific, our findings under this schedule of work aligned with an account of IFG as a buffer for predictions of complex structure – a definition which, in one fell swoop, would account for working memory (Yan et al., 2021; Rogalsky et al., 2008; Smith & Jonides, 1999; Caplan & Waters, 1999), structure-building (Matchin & Hickok, 2020; Hagoort, 2005; Friederici, 2002; Grodzinsky, 2000), and predictive (Matchin et al.,

2017) accounts of function. The formalisation of these notions under the model is therefore a key future goal.

Finally, we proposed extensive roles for an ordinal positional code within VS-BIND, providing evidence that this is likely a boundary-relative positional code that appears to be used in multiple regions including dorsolateral prefrontal cortex and the hippocampus (Carpenter et al., 2018; Shimbo et al., 2021). However, one criticism of an ordinal positional code might be that it does not appear to support transitive inference – specifically, the ability to infer relative positions between novel pairs of referents based on prior exposure to the relative locations of a subset of pairs only (Barron, Reeve, et al., 2020). This ability is especially relevant to the mapping of spatial domains despite limited traversal of the environment. Transitive inference is increasingly considered to be the preserve of the hippocampus, in both spatial and nonspatial domains (Barron, Reeve, et al., 2020; Zeithamova et al., 2012). Thus, transitive inference is important to consider in accounts of item positional encoding, especially with respect to codes hypothesised to support hippocampal function. However, both within our VSA model, and within other proposed VSA encodings of cognitive structures (Lu et al., 2019), it has been suggested that one possible method of generating positional codes is repeated (fractional) binding of a positional basis vector (Plate, 1992). This scheme generates a range of positional codes, but additionally results in the unique property that the binding of an existing "absolute" position (representing the position of some item, A) with a new "offset" position (representing the comparison of the position of another item B, with the position of A) results in a new positional code equal to the intuited output position (i.e., the position of B). This is positional transitive inference. The positional code we posited is therefore compatible with such notions, which raises exciting possibilities for the broader applications of the representational schemes we described and future modelling of hippocampal function using the VS-BIND approach.

In general, we anticipate an increasing role for large-scale, integrative models of neurocomputational mechanisms. VS-BIND is one part of a wider move towards large-scale simulation and hypothesis testing. With this in mind, a useful future goal might be to curate an atlas of neurocomputational hypotheses or "brain computations", leveraging the parsimony, flexibility and intepretability of vector-symbolic accounts to produce a formal documentation of region-specific functional roles. A similar concept has already been enacted, using unconstrained

mathematical operations, in the form of MRC-CBSU/Cambridge University's "Kymata Atlas", an atlas of site-specific hypotheses of brain function (https://www.nitrc.org/projects/kymata_atlas/).

6.1.4 The representational dynamics of auditory sequence processing

Our novel causality analysis ("representational Granger causality"; Chapter 5) revealed thought-provoking findings that were potentially consistent with notions of predictive coding in the frontotemporal language network (see above). However, due to the low number of conditions, and the number of trials that needed to be aggregated into each condition in order to undertake analysis on data with a high signal-to-noise ratio, we were unable to conclusively separate all putative signals from one another. Specifically, we posit that our "presented A" potentially contains "predicted B" information, which in itself, when compared to the "per-syllabic" flow networks, provided promising insights. However, ideally, we would perform this analysis on a wider set of conditions in order to parametrically vary key aspects of sequence structure with greater granularity than "adjacent/non-adjacent" and "A1/A2", for example. In future, we aim to undertake more extensive parameterisation of sequence structure to reveal conclusive effects associated with the alteration of sequence length and identity. Based on our model and the findings of Carpenter et al. (2018), for example, alterations of sequence length would have the potential to reveal putative "boundary-relative" positional codes, whilst increased numbers of item identities and parametric variation of the permissible **A→B** pairs could enable the partial separation of "presented" and "predicted" representations under the representational Granger causal analysis, providing stronger evidence for specific predictive codes in implicated regions.

6.2 Conclusion

The goals of this schedule of research were to characterise human abilities to concurrently learn adjacent and non-adjacent sequencing dependencies within a naturalistic artificial grammar learning paradigm, the *AxB task*; to characterise electrophysiological responses under the same task and establish correspondence with existing findings; to hone a neurocomputational account of sequence processing, in order to inform our understanding of critical encodings and mechanisms; and to develop novel methods to support model-data comparisons. The studies reported in this thesis show that healthy adults can learn a flexible adjacent/non-adjacent dependency relationship, revealing implicit sensitivity to the grammar even in the absence of explicit performance effects. Behavioural and intracranial electrode obtained on the task in neurosurgical patients revealed that, despite chance behavioural performance, patients exhibited electrophysiological responsiveness to sequence grammaticality of selected sequences, including a canonically syntactic LAN-like response observed in vlPFC following presentation of adjacent dependencies. A novel neurocomputational model, VS-BIND, triangulated evidence clarifying putative mechanisms in the fronto-temporal language network. Subsequent multivariate analysis of the intracranial data was consistent with the presence of ordinal encodings in vlPFC and precentral gyrus, as specified in VS-BIND. Novel causal analysis also tentatively suggested top-down transmission of syntactic predictions during the AG task, largely in the opposite direction to stimulus encodings, consistent with predictive coding accounts, and suggested potential roles for the temporoparietal junction and frontal operculum during grammaticality processing. These results and novel analyses have unique potential to inform neurocomputational accounts of language and cognition.

Statement of originality

I hereby declare that the work presented in this submission is my own and to the best of my knowledge contains no materials or intellectual content previously published or written by another person, or substantial portions of content previously accepted for the award of any other degree or diploma at Newcastle University or any other degree-granting institution, except where duly acknowledged.

Ryan Michael Calmus

July 2021

Appendix 1: Additional work by the author

Calmus, R. (2021) Sequence structure in the human brain: Vector symbolic architectures as a neurocomputational lingua franca. *VSAONLINE Winter 2021 Webinar Series.* Online [https://sites.google.com/ltu.se/vsaonline/winter-2021].

Calmus, R., Wilson, B., Kikuchi, Y., Kocsis, Z., Kawasaki, H., Griffiths, T., Howard, M. A. III & Petkov, C. (2019) Neurocomputational transformations of auditory sequences into cognitive structures. *Advances and Perspectives in Auditory Neuroscience 2020*. Online. #88.

Learoyd, A.E., **Calmus, R.,** Cunningham, C.N., England, T.J., Farr, T.D., Fone, K.C.F., Kendall, D.A., O'Sullivan, S.E. & Trueman, R.C. (2021) A pilot of the feasibility and usefulness of an aged obese model for use in stroke research. *Wellcome Open Research*. 6.

Kocsis, Z., Jenison, R. L., McMurray, B., Rhone, A. E., Sarrett, M. E., Gander, P. E., Nourski, K. V., Steinschneider, M., **Calmus, R. M.,** Kawasaki, H., Greenlee, J. D., Kovach, C. K., Griffiths, T. D., Howard, M. A. III & Petkov, C. (2019) Anterior temporal lobe disconnection disrupts auditory cortical oscillatory neural responses to speech in the human brain. *Society for Neuroscience 2019*. Chigaco, IL. 138.01 / J14.

Kikuchi, Y., Kovach, C. K., **Calmus, R.,** Gander, P. E., Rhone, A. E., Nourski, K. V., Kawasaki, H., Griffiths, T. D., Howard, M. A. III & Petkov, C. (2018) Predictive auditory sequence learning modulates inter-regional oscillatory coupling in human intracranial recordings. *Advances and Perspectives in Auditory Neuroscience 2018*. San Diego, CA. #13.

Appendix 2: Supplementary figures

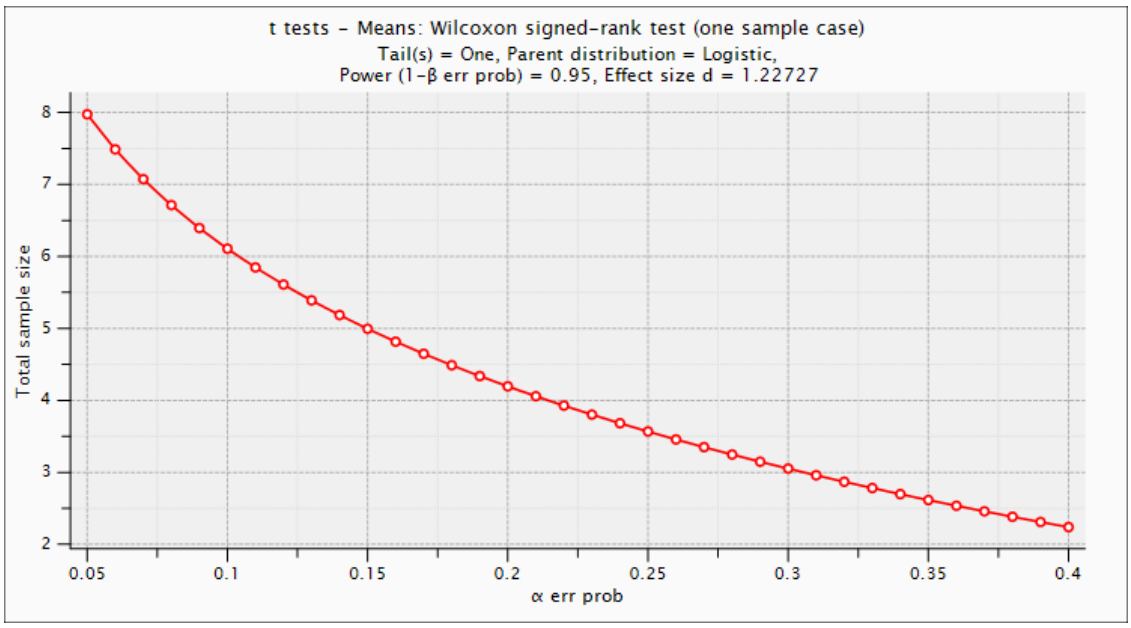


Figure 0.1: **Required minimum sample size to reach one-tailed group significance versus chance performance in a future study.** Making the crude assumption of a logistic parent distribution, whose tails are a marginally better fit for the behaviour of our bimodal final run performance data than a single normal distribution, for $\alpha = .05$, it can be seen that the required power is reached using a sample size of 8 participants. Assuming a normal parent distribution (not shown), the suggested sample size is still between 8 and 9 participants. ***See Chapter 3*** for context.

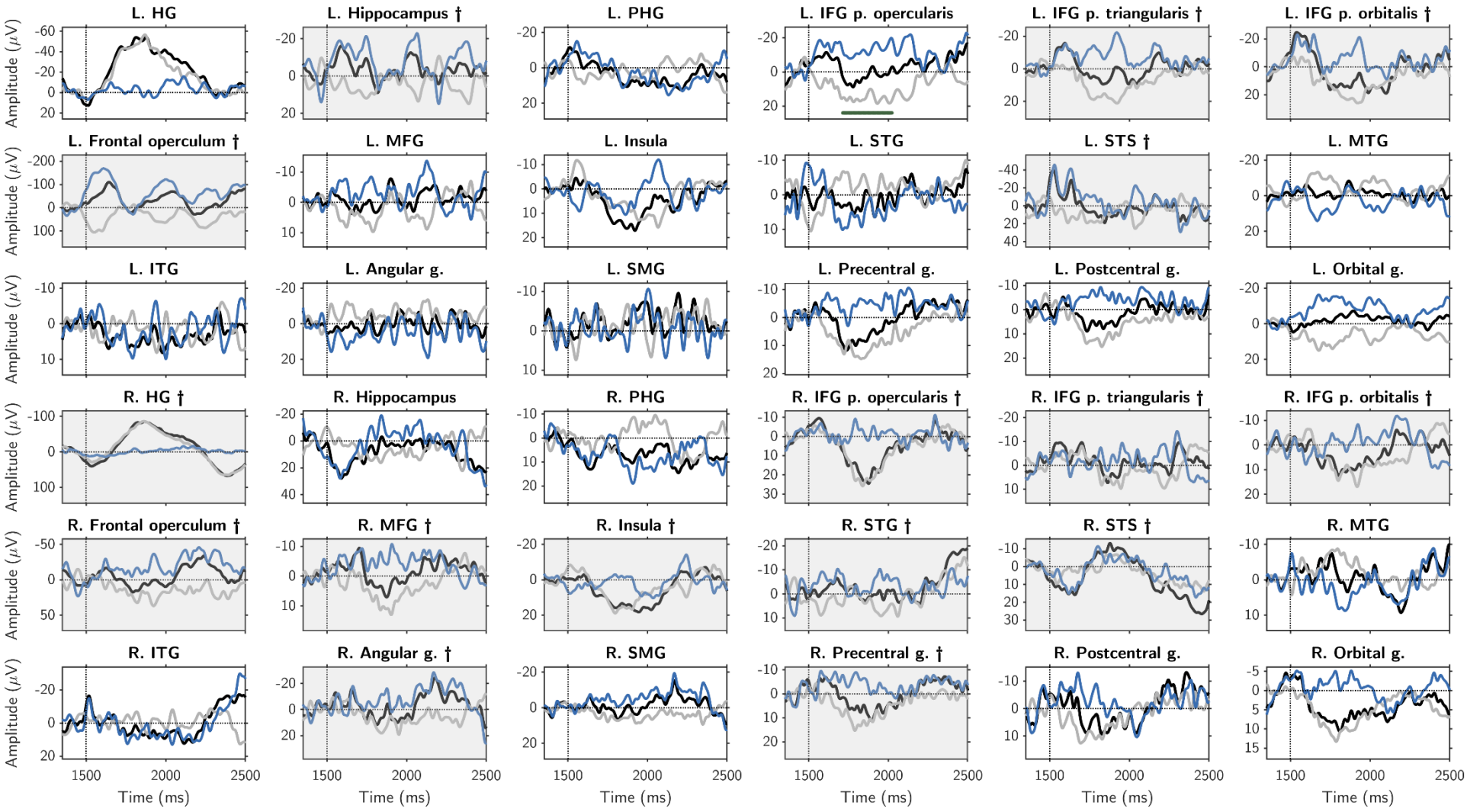


Figure 0.2: **Cohort (n = 12) ERPs, adjacent grammaticality contrast (violation – consistent, adjacent only, bilateral ROIs). The left hemispheric results were previously shown in Figure 3.8. Both hemispheres shown here for completeness.** ***See Chapter 3 for context.***

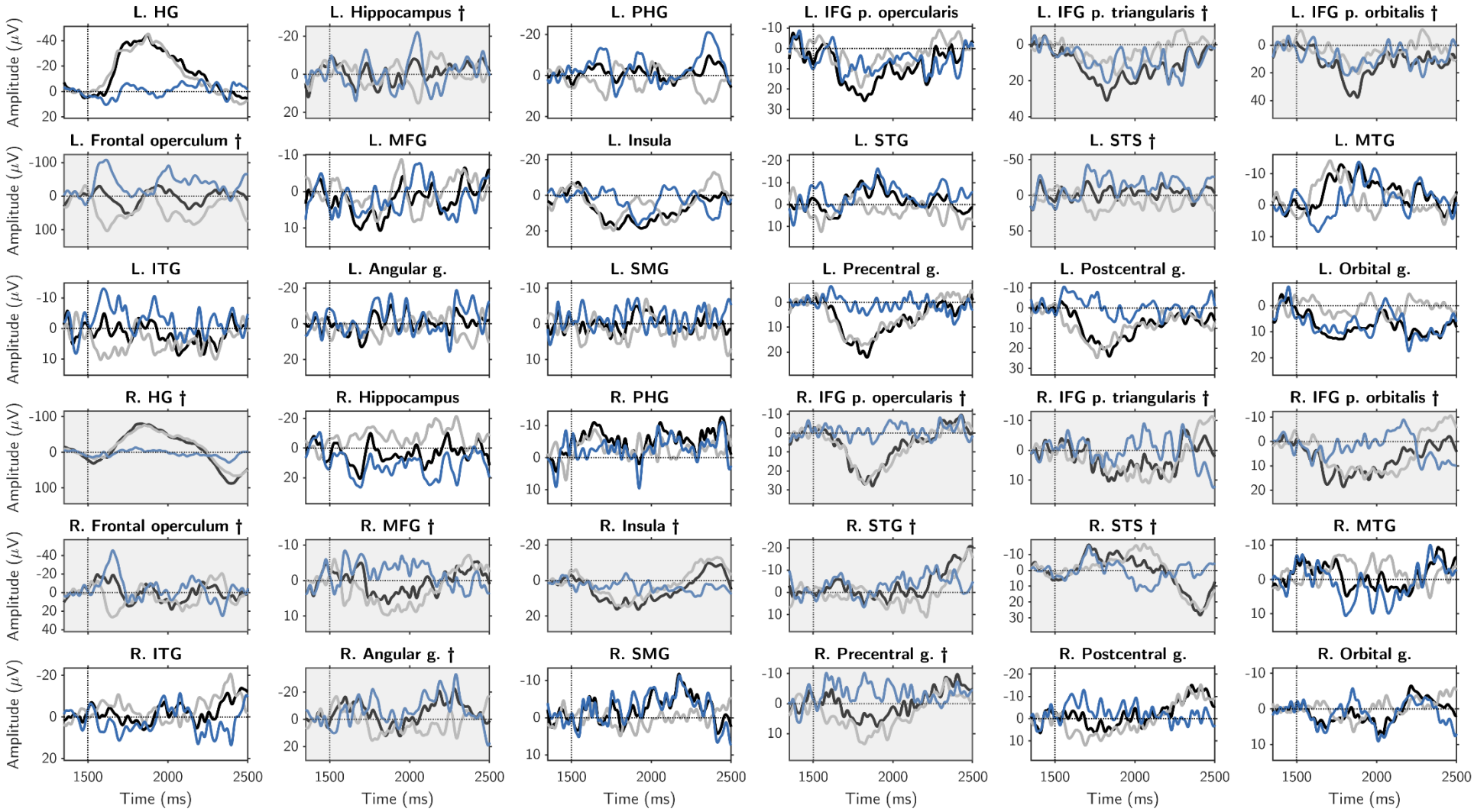


Figure 0.3: **Cohort (n = 12) ERPs, non-adjacent grammaticality contrast (violation – consistent, non-adjacent only, bilateral ROIs). Left hemispheric results previously shown in Figure 3.8. Bilateral results shown here for completeness.** ***See Chapter 3 for context.***

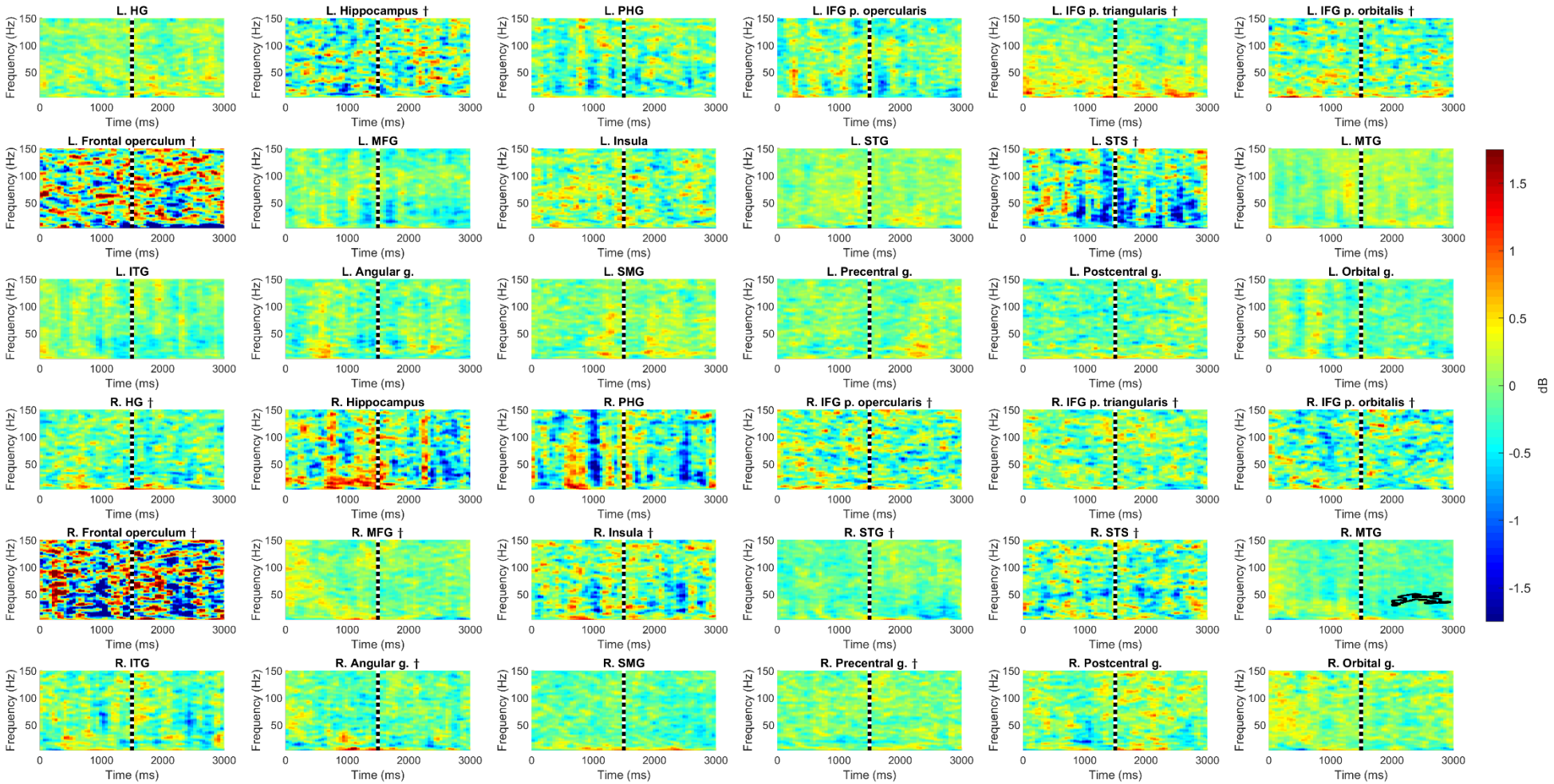


Figure 0.4: **ERSP contrast (violation minus consistent), pooling electrodes at cohort level (n = 12 participants).** ERSP differences (0–3000 ms, 3–150 Hz) are shown as colours on a scale from -1.75 to 1.75 dB. Black contours surround areas of significance as determined by permutation testing ($\alpha = .05$, $\ll 10\,000$ replicates, cluster-corrected for multiple comparisons; only significant cluster in R. MTG). ***See Chapter 3 for context.***

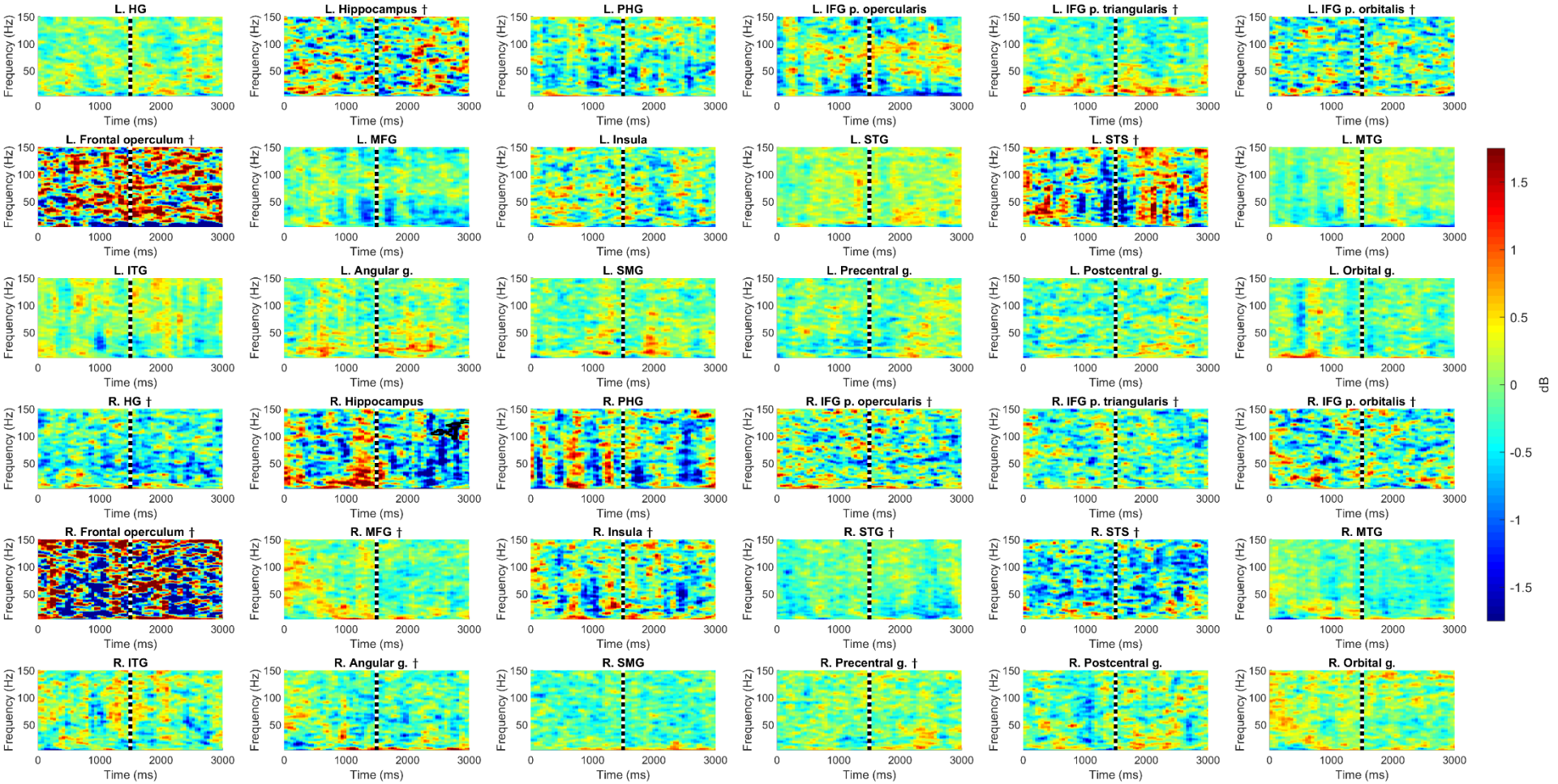


Figure 0.5: **ERSP contrast (violation minus consistent, adjacent only), pooling electrodes at cohort level (n = 12 participants).** ERSP differences (0–3000 ms, 3–150 Hz) are shown as colours on a scale from -1.75 to 1.75 dB. Black contours surround areas of significance as determined by permutation testing ($\alpha = .05$, $\ll$10 000 replicates, cluster-corrected for multiple comparisons; only significant cluster in R. Hippocampus). ***See Chapter 3 for context.***

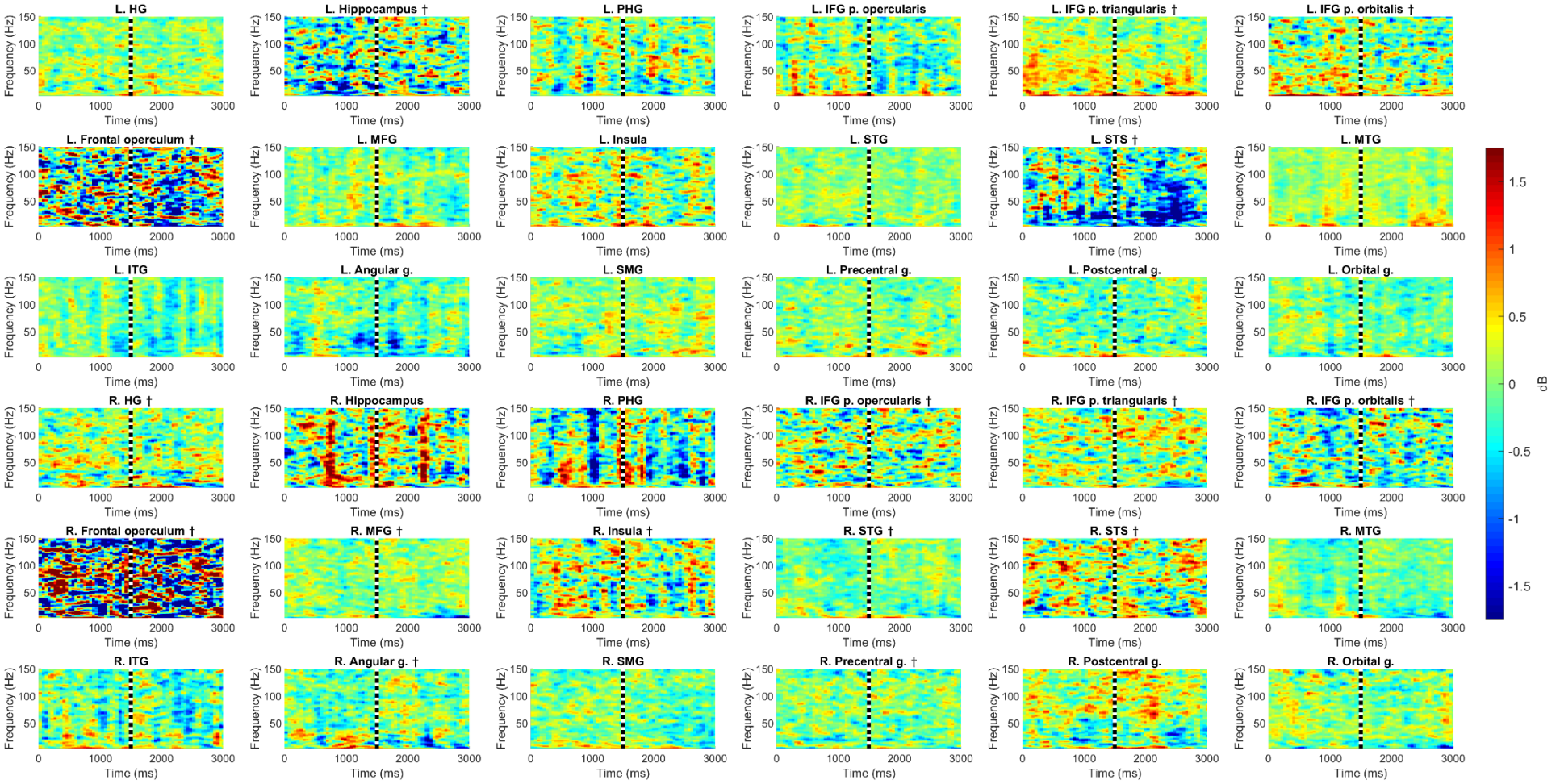


Figure 0.6: **ERSP contrast (violation minus consistent, non-adjacent only), pooling electrodes at cohort level (n = 12 participants).** ERSP differences (0–3000 ms, 3–150 Hz) are shown as colours on a scale from -1.75 to 1.75 dB. Black contours surround areas of significance as determined by permutation testing ($\alpha = .05$, $\ll$10 000 replicates, cluster-corrected for multiple comparisons; no significant clusters). ***See Chapter 3 for context.***

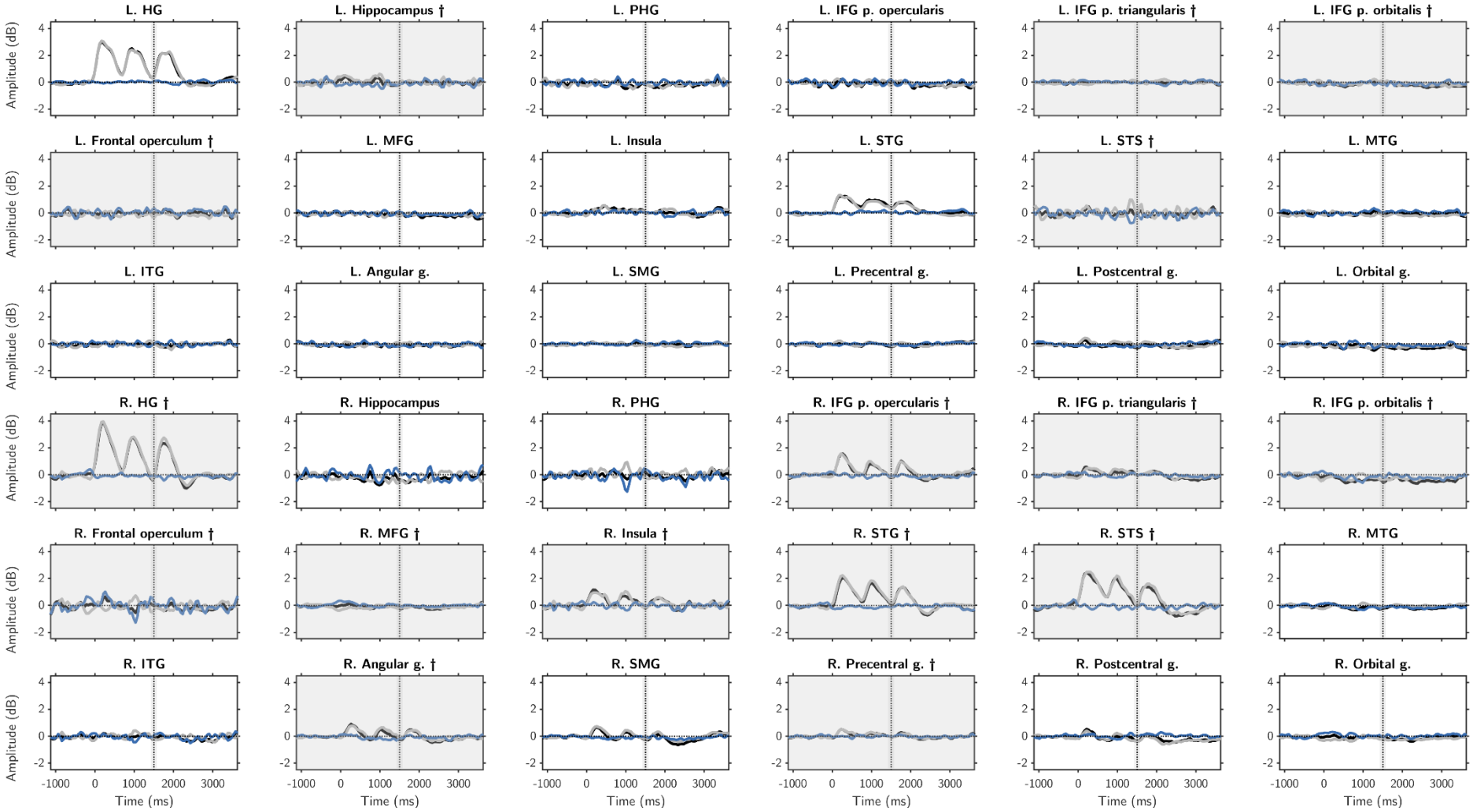


Figure 0.7 **High-gamma power (HGP) contrast (violation minus consistent, n = 12 participants).** Black trace: violation; grey trace: consistent; blue trace: difference. Green points denote significance as determined by permutation testing ($\alpha = .05$, $\ll 10\,000$ replicates, cluster-corrected; no significant clusters). Grey plots marked with a dagger (†) have insufficient coverage to reach α. ***See Chapter 3 for context.***

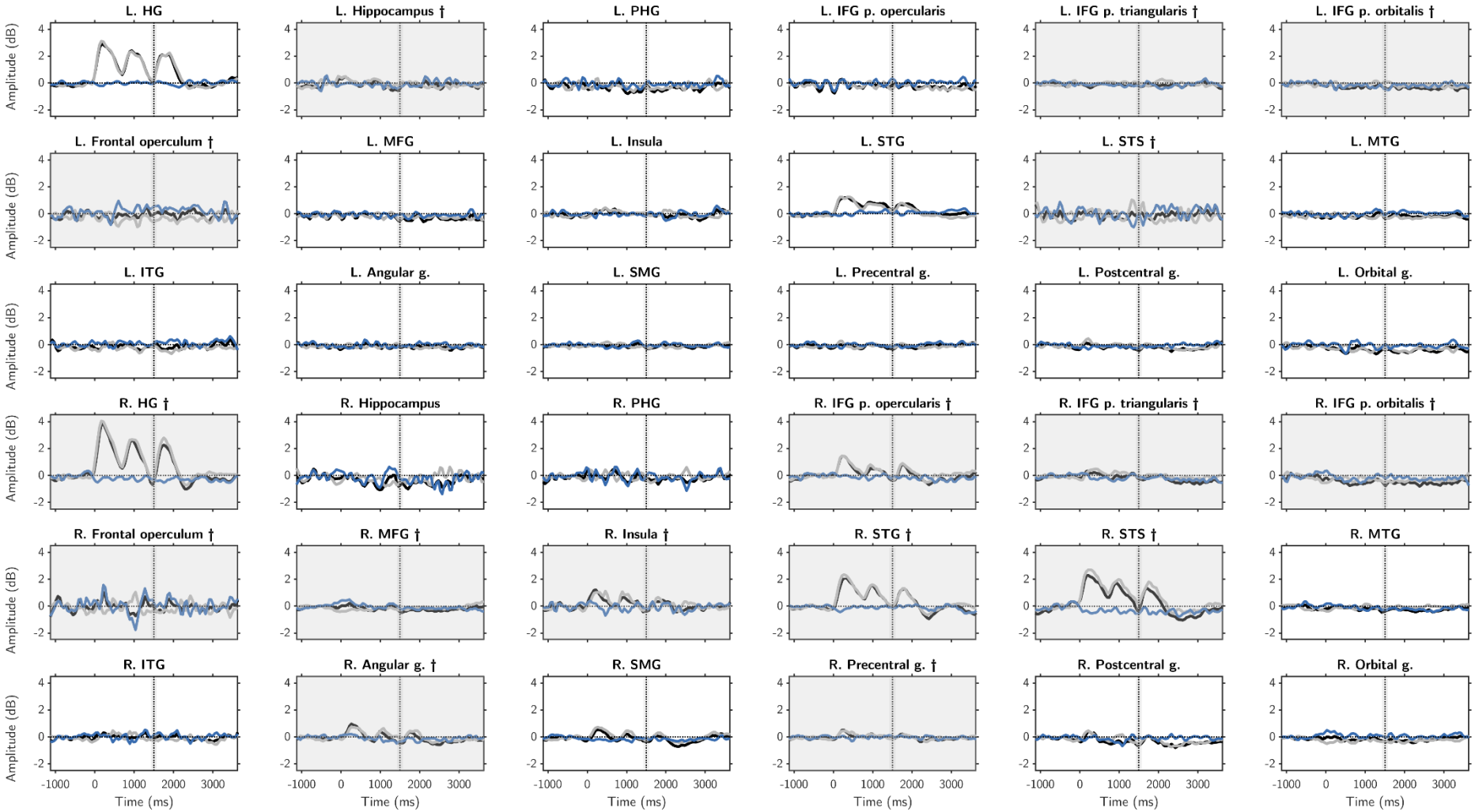


Figure 0.8: **High-gamma power (HGP) contrast (*adjacent* violation minus consistent, n = 12 participants).** Black trace: violation; grey trace: consistent; blue trace: difference. Green points denote significance as determined by permutation testing (α = .05, $\ll$10 000 replicates, cluster-corrected; no significant clusters). Grey plots marked with a dagger (†) have insufficient coverage to reach α. ***See Chapter 3 for context.***

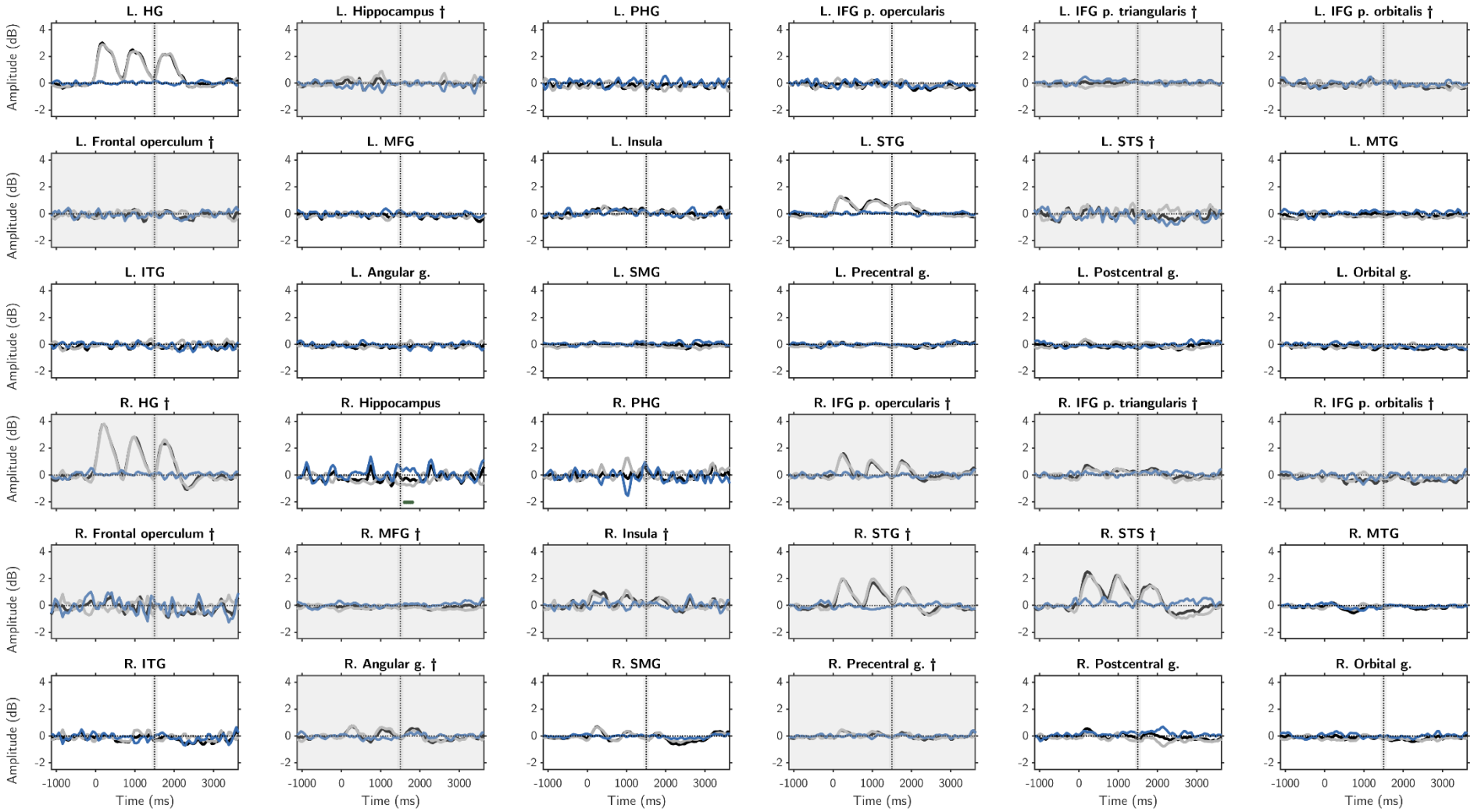


Figure 0.9: **High-gamma power (HGP) contrast (*non-adjacent* violation minus consistent, n = 12 participants).** Black trace: violation; grey trace: consistent; blue trace: difference. Green points denote significance as determined by permutation testing ($\alpha = .05$, $\ll 10\,000$ replicates, cluster-corrected). See R. Hippocampus. Grey plots marked with a dagger (†) have insufficient coverage to reach α. ***See Chapter 3 for context.***

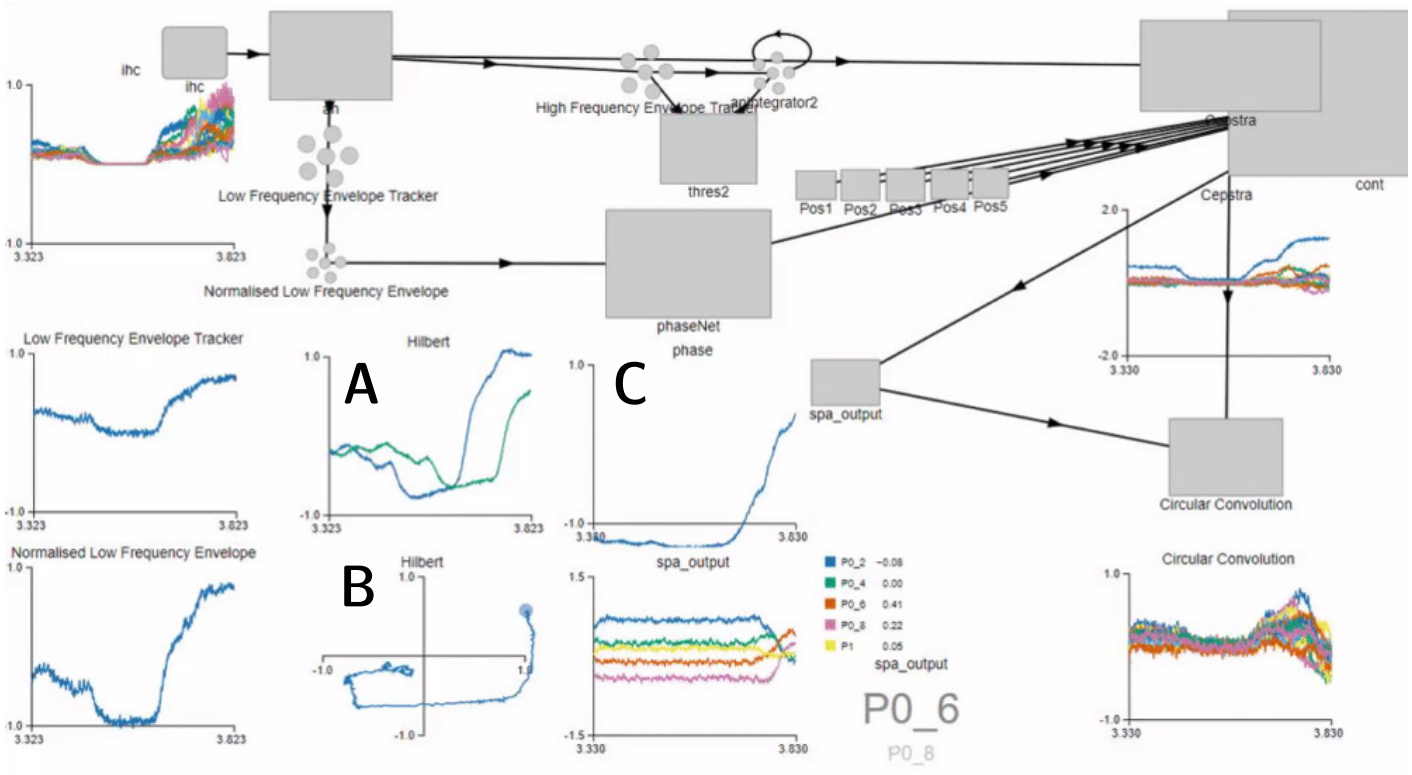


Figure 0.10: **Screenshot of spiking neural simulation of stimulus-driven speech segmentation model.** Spiking neural simulation was accomplished using Nengo (Bekolay et al., 2014) and the screenshot shows a single frame of a browser-based dynamic simulation. Blocks in a directed graph (roughly, top right diagonal half) depict connected neural ensembles each containing hundreds of leaky-integrate-and-fire (LIF) neurons. The bottom left line plots shows the representations instantiated by some of these populations. Of especial importance here are the "Hilbert" line plot and Argand diagrams (**A** and **B**), which both show the same information. A neural representation of the low-frequency speech envelope is delayed using a biologically plausible delay filter to produce two staggered signals ("Hilbert" line plot, **A**, showing original and delayed envelopes). These signals form real and imaginary components of a complex number (see Argand diagram, **B**), from which the phase angle can be readily extracted to produce a neural estimate of instantaneous phase (**C**), as used subsequently within the vector symbolic speech segmentation model; ***see Chapter 4***.

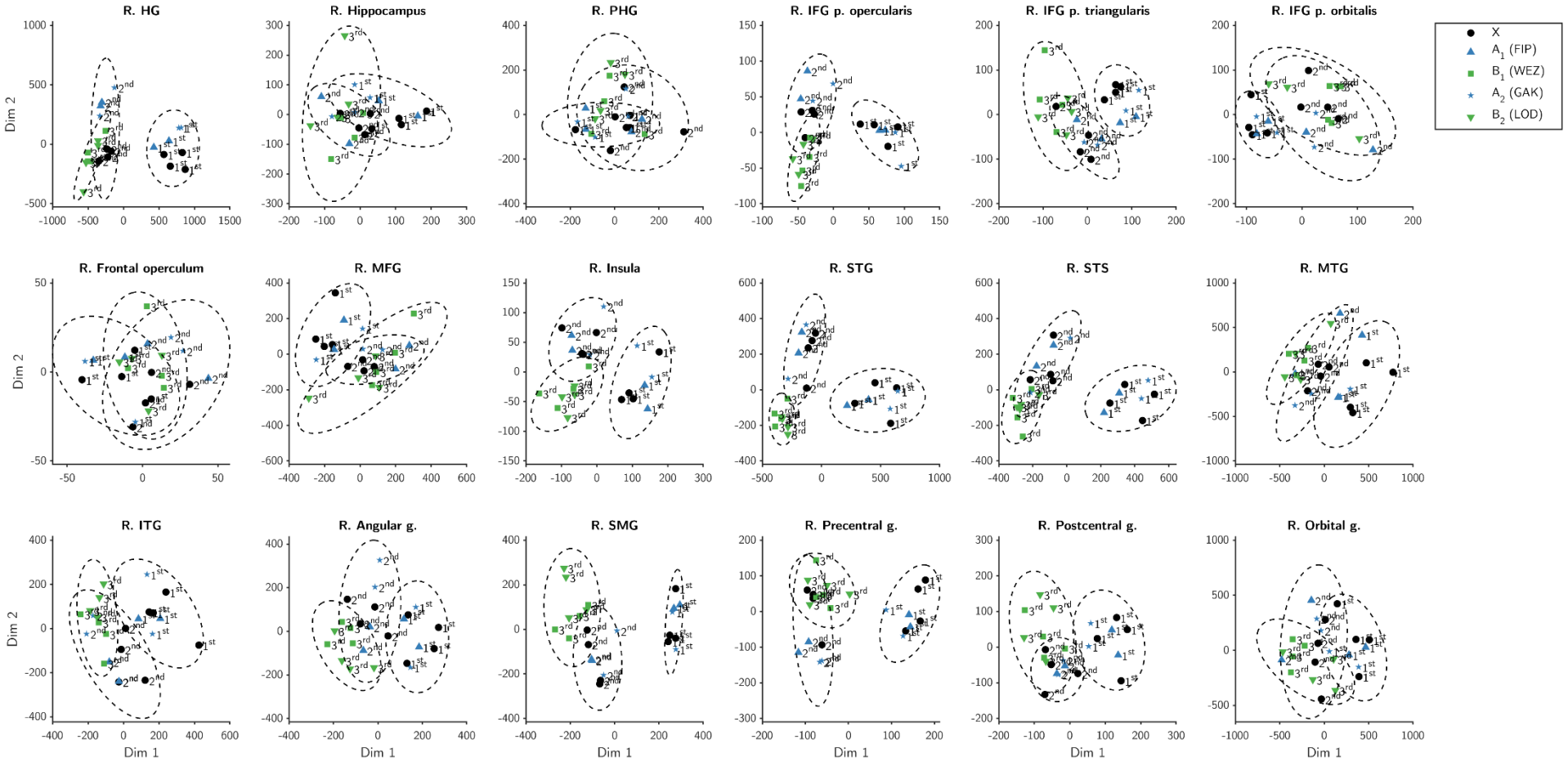


Figure 0.11: **Multidimensional scaling of Euclidean RDM comparing responses to syllables over time (n = 12 participants, right hemisphere).** Dashed ellipses depict 95% confidence boundaries around each cluster of identically positioned syllables. It can be seen that, unlike the left hemisphere (see Figure 5.6), the only region that shows a clean separation potentially on the basis of ordinality is superior temporal gyrus (STG). However, this follows a pattern of markedly shrinking confidence ellipses that hints at repetition suppression effects as a possible driver of this representational geometry. ***See Chapter 5 for context.***

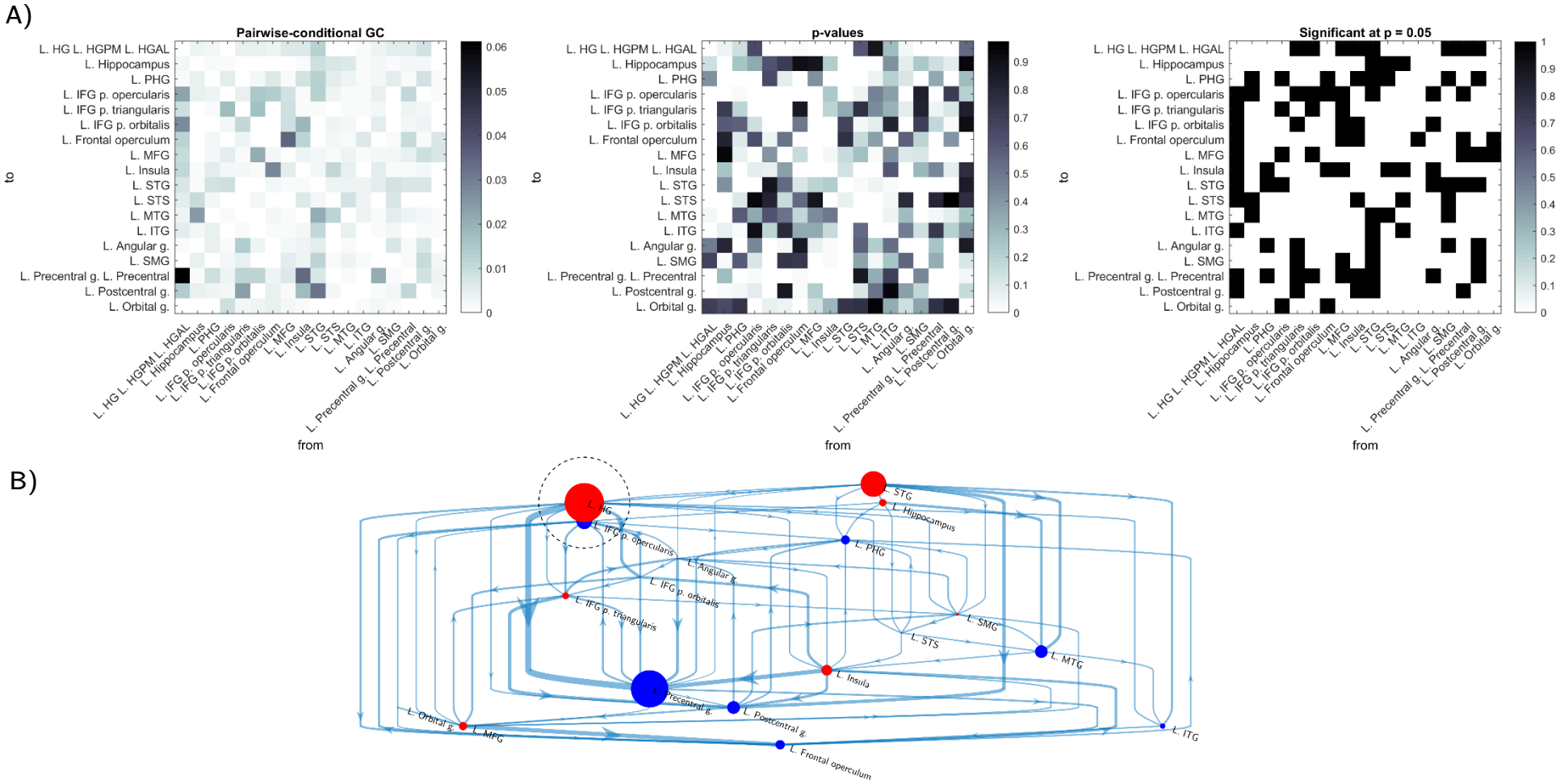


Figure 0.12: **Granger causality results computed on ECoG data.** Data was recorded at 2000 Hz, SVD-denoised and downsampled to 500 Hz during preprocessing. Data was then epoched, averaged over trials and downsampled again by a factor of 12.5 to generate a 40 Hz signal, the sampling rate used in the representational Granger analysis. Analysis was performed using the MVGC toolbox. Significant causal associations were found between many regions (panel A, right), connecting many ROIs. This produced a densely connected causal flow graph (panel B). ***See Chapter 5.***

Appendix 3: Behavioural questionnaire

Confidential Study Questionnaire

You have the option to omit any responses, but note that all given information here is completely anonymous and your answers will help us to better analyze your experimental data.

Experimenter:_____
Study: _____

Anonymous Code (given by the Study Coordinator): _____
Date: _____

Your Gender: (circle one) M F Other Prefer not to say
Your Age: _____
Are you Right or Left handed: (circle one) Right Left Both

Would you consider yourself bilingual? (Yes or No) _____
If Yes, please describe your level of fluency in your second language.
_____
_____

Do you currently suffer from a cold, flu, or ear infection? (circle) Y N

Is there anything impairing your hearing or sight right now? (circle) Y N
If Yes, please describe: _____

Has your hearing been normal, as far as you can tell? Y N
If No, describe (for example, I had an ear infection in 1995 but my hearing is normal now).
_____
_____

Has your vision been normal as far as you know? Y N
If No, describe (for example, I had blurry vision during the World Cup in 2006 but it is normal again).
_____
_____

Do you wear glasses or contact lenses? Y N
If Yes, are you wearing them today?
_____
_____

Do you have a Language or Learning impairment (such as dyslexia)? Y N

Do you frequently listen to loud music or were you exposed to loud sounds (like machines, construction work, or frequent loud iPod music listening)? Y N

If Yes, describe including what type of sounds:

_____

_____

How many hours of sleep approximately did you get last night? Please circle.

Less than 2 hours	2-4 hours	4-6 hours	6-8 hours	8-10 hours	More than 10 hours
1	2	3	4	5	6

How tired do you feel right now? Please tick one of the options that best describes you.

Very tired	Tired	Slightly tired	Rested	Energised
1	2	3	4	5

How well do you feel that you learnt the structure of the sequences?

Very Well	Quite Well	Moderately	A little	Not at all
1	2	3	4	5

Please try and list any rules you noticed about the sequence:

_____

_____

_____

Bibliography

Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A. & Arshad, H. (2018) State-of-the-art in artificial neural network applications: A survey. *Heliyon*. 4 (11), e00938.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H. & Merzenich, M.M. (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*. 98 (23), 13367–13372.

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 19 (6), 716–723.

Alamia, A., Gauducheau, V., Paisios, D. & VanRullen, R. (2020) Comparing feedforward and recurrent neural network architectures with human behavior in artificial grammar learning. *Scientific Reports*. 10 (1), 22172.

Alemán Bañón, J. & Rothman, J. (2019) Being a Participant Matters: Event-Related Potentials Show That Markedness Modulates Person Agreement in Spanish. *Frontiers in Psychology*. 10, 746.

Alfonso-Reese, L.A., Ashby, F.G. & Brainard, D.H. (2002) What makes a categorization task difficult? *Perception & Psychophysics*. 64 (4), 570–583.

Alhama, R.G. & Zuidema, W. (2019) A review of computational models of basic rule learning: The neural-symbolic debate and beyond. *Psychonomic Bulletin & Review*. 26 (4), 1174–1194.

Anderson, C.W., Knight, J.N., O'Connor, T., Kirby, M.J. & Sokolov, A. (2006) Geometric subspace methods and time-delay embedding for EEG artifact removal and classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 14 (2), 142–146.

Angrilli, A., Penolazzi, B., Vespignani, F., De Vincenzi, M., Job, R., Ciccarelli, L., Palomba, D. & Stegagno, L. (2002) Cortical brain responses to semantic incongruity and syntactic violation in Italian language: an event-related potential study. *Neuroscience Letters*. 322 (1), 5–8.

Asano, E., Juhász, C., Shah, A., Muzik, O., Chugani, D.C., Shah, J., Sood, S. & Chugani, H.T. (2005) Origin and Propagation of Epileptic Spasms Delineated on Electrocorticography. *Epilepsia*. 46 (7), 1086–1097.

Aslin, R.N., Saffran, J.R. & Newport, E.L. (1998) Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*. 9 (4), 321–324.

Attaheri, A., Kikuchi, Y., Milne, A.E., Wilson, B., Alter, K. & Petkov, C.I. (2015) EEG potentials associated with artificial grammar learning in the primate brain. *Brain and Language*. 148, 74–80.

Aubin, S., Voelker, A.R. & Eliasmith, C. (2017) Improving With Practice: A Neural Model of Mathematical Development. *Topics in Cognitive Science*. 9 (1), 6–20.

Bahlmann, J., Schubotz, R.I., Mueller, J.L., Koester, D. & Friederici, A.D. (2009) Neural circuits of hierarchical visuo-spatial sequence processing. *Brain Research*. 1298, 161–170.

Balezeau, F., Wilson, B., Gallardo, G., Dick, F., Hopkins, W., Anwander, A., Friederici, A.D., Griffiths, T.D. & Petkov, C.I. (2020) Primate auditory prototype in the evolution of the arcuate fasciculus. *Nature Neuroscience*. 23 (5), 611–614.

Barczak, A., O'Connell, M.N., McGinnis, T., Ross, D., Mowery, T., Falchier, A. & Lakatos, P. (2018) Top-down, contextual entrainment of neuronal oscillations in the auditory thalamocortical circuit. *Proceedings of the National Academy of Sciences*. 115 (32), E7605–E7614.

Barnett, L. & Seth, A.K. (2014) The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*. 223, 50–68.

Barron, H.C., Auksztulewicz, R. & Friston, K. (2020) Prediction and memory: A predictive coding account. *Progress in Neurobiology*. 192, 101821.

Barron, H.C., Reeve, H.M., Koolschijn, R.S., Perestenko, P.V., Shpektor, A., Nili, H., Rothaermel, R., Campo-Urriza, N., O'Reilly, J.X., Bannerman, D.M., Behrens, T.E.J. & Dupret, D. (2020) Neuronal Computation Underlying Inferential Reasoning in Humans and Mice. *Cell*. 183 (1), 228-243.e21.

Barth, W., Mutzel, P. & Jünger, M. (2004) Simple and Efficient Bilayer Cross Counting. *Journal of Graph Algorithms and Applications*. 8 (2), 179–194.

Bartlett, M. & Kazakov, D. (2005) The origins of syntax: from navigation to language. *Connection Science*. 17 (3–4), 271–288.

Bastiaansen, M. & Hagoort, P. (2015) Frequency-based Segregation of Syntactic and Semantic Unification during Online Sentence Level Language Comprehension. *Journal of Cognitive Neuroscience*. 27 (11), 2095–2107.

Bastiaansen, M., Magyari, L. & Hagoort, P. (2010) Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience*. 22 (7), 1333–1347.

Batterink, L. & Neville, H.J. (2013) The Human Brain Processes Syntax in the Absence of Conscious Awareness. *Journal of Neuroscience*. 33 (19), 8528–8533.

Batterink, L.J. & Paller, K.A. (2017) Online neural monitoring of statistical learning. *Cortex; a journal devoted to the study of the nervous system and behavior*. 90, 31–45.

Batterink, L.J., Paller, K.A. & Reber, P.J. (2019) Understanding the Neural Bases of Implicit and Statistical Learning. *Topics in Cognitive Science*. 11 (3), 482–503.

Batterink, L.J., Reber, P.J., Neville, H.J. & Paller, K.A. (2015) Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*. 8362–78.

Bayati, M., Neher, T., Melchior, J., Diba, K., Wiskott, L. & Cheng, S. (2018) Storage fidelity for sequence memory in the hippocampal circuit. *PLoS ONE*. 13 (10), e0204685.

Bekolay, T. (2016) *Biologically inspired methods in speech recognition and synthesis: closing the loop*. [Online]. Waterloo, ON, Canada: University of Waterloo.

Bekolay, T., Bergstra, J., Hunsberger, E., DeWolf, T., Stewart, T.C., Rasmussen, D., Choo, X., Voelker, A.R. & Eliasmith, C. (2014) Nengo: A Python tool for building large-scale functional brain models. *Frontiers in Neuroinformatics*. 7 (48).

Bell, A.J. & Sejnowski, T.J. (1997) The "independent components" of natural scenes are edge filters. *Vision Research*. 37 (23), 3327–3338.

Beres, A.M. (2017) Time is of the Essence: A Review of Electroencephalography (EEG) and Event-Related Brain Potentials (ERPs) in Language Research. *Applied Psychophysiology and Biofeedback*. 42 (4), 247–255.

Berry, D.C. & Broadbent, D.E. (1984) On the Relationship between Task Performance and Associated Verbalizable Knowledge. *The Quarterly Journal of Experimental Psychology Section A*. 36 (2), 209–231.

Beyeler, M., Rounds, E.L., Carlson, K.D., Dutt, N. & Krichmar, J.L. (2019) Neural correlates of sparse coding and dimensionality reduction. *PLOS Computational Biology*. 15 (6), e1006908.

Bickerton, D. & Szathmáry, E. (eds.) (2009) *Biological Foundations and Origin of Syntax*. First Edition edition. Cambridge, Mass: The MIT Press.

Binder, J.R. (2015) The Wernicke area. *Neurology*. 85 (24), 2170–2175.

Bitan, T., Cheon, J., Lu, D., Burman, D.D., Gitelman, D.R., Mesulam, M.-M. & Booth, J.R. (2007) Developmental changes in activation and effective connectivity in phonological processing. *NeuroImage*. 38 (3), 564–575.

Blanca, M.J., Alarcón, R., Arnau, J., Bono, R. & Bendayan, R. (2017) Non-normal data: Is ANOVA still a valid option? *Psicothema*. 29 (4), 552–557.

Boersma, P. & Weenink, D. (2016) *Praat: doing Phonetics by Computer* (Version 6.0.20). [Computer software].

Bornkessel-Schlesewsky, I. & Schlesewsky, M. (2019) Toward a Neurobiologically Plausible Model of Language-Related, Negative Event-Related Potentials. *Frontiers in Psychology*. 10, 298.

van den Bos, E., Christiansen, M.H. & Misyak, J.B. (2012) Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language*. 67 (4), 507–520.

Botvinick, M. & Watanabe, T. (2007) From Numerosity to Ordinal Rank: A Gain-Field Model of Serial Order Representation in Cortical Working Memory. *Journal of Neuroscience*. 27 (32), 8636–8642.

Boutros, N.N., Gjini, K., Urbach, H. & Pflieger, M.E. (2011) Mapping repetition suppression of the N100 evoked response to the human cerebral cortex. *Biological psychiatry*. 69 (9), 883–889.

Bowers, J.S. (2009) On the biological plausibility of grandmother cells: implications for neural network theories in psychology and neuroscience. *Psychological Review*. 116 (1), 220–251.

Braaksma, B., Broer, H. & Takens, F. (eds.) (1985) *Dynamical Systems and Bifurcations*. Lecture Notes in Mathematics 1125. Berlin: Springer.

Brandes, U. & Köpf, B. (2002) 'Fast and Simple Horizontal Coordinate Assignment', in Petra Mutzel, Michael Jünger, & Sebastian Leipert (eds.) *Graph Drawing*. Lecture Notes in Computer Science. [Online]. 2002 Berlin, Heidelberg: Springer. pp. 31–44.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.-F., Sommer, J., Wolbers, T. & Knecht, S. (2005) Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage*. 25 (3), 958–968.

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D.J. & Pylkkänen, L. (2012) Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*. 120 (2), 163–173.

Brent, M.R. (1999) Speech segmentation and word discovery: a computational perspective. *Trends in Cognitive Sciences*. 3 (8), 294–301.

Broca, P. (1861) *Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole)*. [Online]. Available from: https://psychclassics.yorku.ca/Broca/aphemie.htm (Accessed 23 June 2021).

Brouwer, H. & Crocker, M.W. (2017) On the Proper Treatment of the N400 and P600 in Language Comprehension. *Frontiers in Psychology*. 8, 1327.

Brouwer, H. & Hoeks, J.C.J. (2013) A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience*. 7, 758.

Buiatti, M., Peña, M. & Dehaene-Lambertz, G. (2009) Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *NeuroImage*. 44 (2), 509–519.

Burton, M.W. & Small, S.L. (2006) Functional Neuroanatomy of Segmenting Speech and Nonspeech. *Cortex*. 42 (4), 644–651.

Burton, M.W., Small, S.L. & Blumstein, S.E. (2000) The role of segmentation in phonological processing: an fMRI investigation. *Journal of Cognitive Neuroscience*. 12 (4), 679–690.

Buzsáki, G. (2002) Theta Oscillations in the Hippocampus. *Neuron*. 33 (3), 325–340.

Buzsáki, G. & Tingley, D. (2018) Space and Time: The Hippocampus as a Sequence Generator. *Trends in Cognitive Sciences*. 22 (10), 853–869.

Buzsáki, G. & Wang, X.-J. (2012) Mechanisms of gamma oscillations. *Annual Review of Neuroscience*. 35, 203–225.

Caffarra, S., Mendoza, M. & Davidson, D. (2019) Is the LAN effect in morphosyntactic processing an ERP artifact? *Brain and Language*. 191, 9–16.

Calmus, R. (2019) *rcalmus/vs-bind: VS-BIND Supplementary Material*. Zenodo.

Calmus, R., Wilson, B., Kikuchi, Y. & Petkov, C.I. (2019) Structured sequence processing and combinatorial binding: neurobiologically and computationally informed hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 375 (1791), 20190304.

Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Kirsch, H.E., Berger, M.S., Barbaro, N.M. & Knight, R.T. (2006) High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex. *Science (New York, N.Y.)*. 313 (5793), 1626–1628.

Caplan, D. & Waters, G.S. (1999) Verbal working memory and sentence comprehension. *The Behavioral and Brain Sciences*. 22 (1), 77–94; discussion 95-126.

Carpenter, A.F., Baud-Bovy, G., Georgopoulos, A.P. & Pellizzer, G. (2018) Encoding of Serial Order in Working Memory: Neuronal Activity in Motor, Premotor, and Prefrontal Cortex during a Memory Scanning Task. *Journal of Neuroscience*. 38 (21), 4912–4933.

ten Cate, C., Gervain, J., Levelt, C.C., Petkov, C.I. & Zuidema, W. (2020) Editors' Review and Introduction: Learning Grammatical Structures: Developmental, Cross-Species, and Computational Approaches. *Topics in Cognitive Science*. 12 (3), 804–814.

Cavanagh, S.E., Hunt, L.T. & Kennerley, S.W. (2020) A Diversity of Intrinsic Timescales Underlie Neural Computations. *Frontiers in Neural Circuits*. 14, 615626.

Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M. & Knight, R.T. (2010) Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*. 13 (11), 1428–1432.

Chen, J., Dastjerdi, M., Foster, B.L., LaRocque, K.F., Rauschecker, A.M., Parvizi, J. & Wagner, A.D. (2013) Human hippocampal increases in low-frequency power during associative prediction violations. *Neuropsychologia*. 51 (12), 2344–2351.

Chen, Y., Shimotake, A., Matsumoto, R., Kunieda, T., Kikuchi, T., Miyamoto, S., Fukuyama, H., Takahashi, R., Ikeda, A. & Lambon Ralph, M.A. (2016) The 'when' and 'where' of semantic coding in the anterior temporal lobe: Temporal representational similarity analysis of electrocorticogram data. *Cortex*. 79, 1–13.

Cheung, V.K.M., Meyer, L., Friederici, A.D. & Koelsch, S. (2018) The right inferior frontal gyrus processes nested non-local dependencies in music. *Scientific Reports*. 8 (1), 3822.

Chomsky, N. (1957) *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1995) *The Minimalist Program*. MIT Press.

Chomsky, N. (1956) Three models for the description of language. *IRE Transactions on Information Theory*. 2 (3), 113–124.

Choo, F.-X. & Eliasmith, C. (2010) 'A Spiking Neuron Model of Serial-Order Recall', in *Proceedings of the Annual Meeting of the Cognitive Science Society*. [Online]. pp. 2188–2193.

Christiansen, M.H. (2019) Implicit Statistical Learning: A Tale of Two Literatures. *Topics in Cognitive Science*. 11 (3), 468–481.

Christiansen, M.H., Conway, C.M. & Onnis, L. (2012) Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes*. 27 (2), 231–256.

Christiansen, M.H., Louise Kelly, M., Shillcock, R.C. & Greenfield, K. (2010) Impaired artificial grammar learning in agrammatism. *Cognition*. 116 (3), 382–393.

Christiansen, M.H. & MacDonald, M.C. (2009) A usage-based approach to recursion in sentence processing. *Language Learning*. 59, 126–161.

Cleeremans, A. & McClelland, J.L. (1991) Learning the structure of event sequences. *Journal of Experimental Psychology: General*. 120 (3), 235–253.

Clos, M., Amunts, K., Laird, A.R., Fox, P.T. & Eickhoff, S.B. (2013) Tackling the multifunctional nature of Broca's region meta-analytically: Co-activation-based parcellation of area 44. *NeuroImage*. 83, 174–188.

Cock, J. (2010) Completely ruled out? Using response time data to investigate biconditional artificial grammar learning. *European Journal of Cognitive Psychology*. 17 (6), 770–784.

Cohen, M.X. (2014) *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press.

Comins, J.A. & Gentner, T.Q. (2010) Working memory for patterned sequences of auditory objects in a songbird. *Cognition*. 117 (1), 38–53.

Cona, G. & Semenza, C. (2017) Supplementary motor area as key structure for domain-general sequence processing: A unified account. *Neuroscience & Biobehavioral Reviews*. 72, 28–42.

Conway, C.M., Bauernschmidt, A., Huang, S.S. & Pisoni, D.B. (2010) Implicit statistical learning in language processing: word predictability is the key. *Cognition*. 114 (3), 356–371.

Conway, C.M. & Christiansen, M.H. (2006) Statistical Learning Within and Between Modalities: Pitting Abstract Against Stimulus-Specific Representations. *Psychological Science*. 17 (10), 905–912.

Conway, C.M. & Pisoni, D.B. (2008) Neurocognitive Basis of Implicit Learning of Sequential Structure and Its Relation to Language Processing. *Annals of the New York Academy of Sciences*. 1145 (1), 113–131.

Cope, T.E., Wilson, B., Robson, H., Drinkall, R., Dean, L., Grube, M., Jones, P.S., Patterson, K., Griffiths, T.D., Rowe, J.B. & Petkov, C.I. (2017) Artificial grammar learning in vascular and progressive non-fluent aphasias. *Neuropsychologia*. 104, 201–213.

Covington, N.V., Brown-Schmidt, S. & Duff, M.C. (2018) The Necessity of the Hippocampus for Statistical Learning. *Journal of Cognitive Neuroscience*. 30 (5), 680–697.

Cox, D.D. & Savoy, R.L. (2003) Functional magnetic resonance imaging (fMRI) 'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*. 19 (2 Pt 1), 261–270.

Crawford, E., Gingerich, M. & Eliasmith, C. (2016) Biologically Plausible, Human-Scale Knowledge Representation. *Cognitive Science*. 40 (4), 782–821.

Cross, Z.R., Zou-Williams, L., Wilkinson, E.M., Schlesewsky, M. & Bornkessel-Schlesewsky, I. (2020) Mini Pinyin: A modified miniature language for studying language learning and incremental sentence processing. *Behavior Research Methods*. 53 (3), 1218–1239.

Cui, Y., Ahmad, S. & Hawkins, J. (2016) Continuous Online Sequence Learning with an Unsupervised Neural Network Model. *Neural Computation*. 28 (11), 2474–2504.

Cumming, G. (1984) Sleep promotion, hospital practice and recovery from illness. *Medical Hypotheses*. 15 (1), 31–37.

Davachi, L. & DuBrow, S. (2015) How the hippocampus preserves order: the role of prediction and context. *Trends in Cognitive Sciences*. 19 (2), 92–99.

Davidson, D.J. & Indefrey, P. (2007) An inverse relation between event-related and time–frequency violation responses in sentence processing. *Brain Research*. 1158, 81–92.

Davis, P.A. (1939) Effects of acoustic stimuli on the waking human brain. *Journal of Neurophysiology*. 2 (6), 494–499.

Day, W.H. & Edelsbrunner, H. (1984) Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of Classification*. 1 (1), 7–24.

Deco, G., Cruzat, J. & Kringelbach, M.L. (2019) Brain songs framework used for discovering the relevant timescale of the human brain. *Nature Communications*. 10 (1), 583.

Dehaene, S. & Changeux, J.-P. (1997) A hierarchical neuronal network for planning behavior. *Proceedings of the National Academy of Sciences*. 94 (24), 13293–13298.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L. & Pallier, C. (2015) The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron*. 88 (1), 2–19.

Dehaene, S., Molko, N., Cohen, L. & Wilson, A.J. (2004) Arithmetic and the brain. *Current Opinion in Neurobiology*. 14 (2), 218–224.

DeKeyser, R. (2003) 'Implicit and Explicit Learning', in *The Handbook of Second Language Acquisition*. [Online]. John Wiley & Sons, Ltd. pp. 312–348.

Delogu, F., Brouwer, H. & Crocker, M.W. (2019) Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition*. 135, 103569.

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 (1), 9–21.

Deocampo, J., King, T. & Conway, C. (2019) Concurrent Learning of Adjacent and Nonadjacent Dependencies in Visuo-Spatial and Visuo-Verbal Sequences. *Frontiers in Psychology*. 10, 1107.

Destrieux, C., Fischl, B., Dale, A. & Halgren, E. (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*. 53 (1), 1–15.

Destrieux, C., Terrier, L.M., Andersson, F., Love, S.A., Cottier, J.-P., Duvernoy, H., Velut, S., Janot, K. & Zemmoura, I. (2017) A practical guide for the identification of major sulcogyral structures of the human cortex. *Brain Structure & Function*. 222 (4), 2001–2015.

Diedrichsen, J. & Kriegeskorte, N. (2017) Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLOS Computational Biology*. 13 (4), e1005508.

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

Doppelmayr, M., Klimesch, W., Pachinger, T. & Ripper, B. (1998) Individual differences in brain dynamics: important implications for the calculation of event-related band power. *Biological Cybernetics*. 79 (1), 49–57.

Doumas, L.A.A. & Hummel, J.E. (2005) 'Approaches to Modeling Human Mental Representations: What Works, What Doesn't, and Why', in *The Cambridge handbook of thinking and reasoning*. [Online]. New York, NY, US: Cambridge University Press. pp. 73–91.

Doumas, L.A.A., Hummel, J.E. & Sandhofer, C.M. (2008) A theory of the discovery and predication of relational concepts. *Psychological Review*. 115 (1), 1–43.

Duff, M. & Brown-Schmidt, S. (2012) The hippocampus and the flexible use and processing of language. *Frontiers in Human Neuroscience*. 6, 69.

Duggento, A., Passamonti, L., Valenza, G., Barbieri, R., Guerrisi, M. & Toschi, N. (2018) Multivariate Granger causality unveils directed parietal to prefrontal cortex connectivity during task-free MRI. *Scientific Reports*. 8 (1), 5571.

Dumont, N.S.-Y. & Eliasmith, C. (2020) 'Accurate representation for spatial cognition using grid cells', in *Proceedings of the Cognitive Science Society*. [Online]. pp. 2367–2373.

Durrant, S.J., Cairney, S.A. & Lewis, P.A. (2016) Cross-modal transfer of statistical information benefits from sleep. *Cortex*. 78, 85–99.

Edelman, S. (2017) Language and other complex behaviors: Unifying characteristics, computational models, neural mechanisms. *Language Sciences*. 62, 91–123.

Eichenbaum, H. (2013) Memory on time. *Trends in cognitive sciences*. 17 (2), 81–88.

Ekstrom, A.D., Spiers, H.J., Bohbot, V.D. & Rosenbaum, R.S. (2018) *Human Spatial Navigation*. Princeton, New Jersey: Princeton University Press.

Eliasmith, C. (2013) *How to Build a Brain: A Neural Architecture for Biological Cognition*. Oxford University Press.

Eliasmith, C., Anderson, C.H., Sejnowski, T.J. & Poggio, T.A. (2004) *Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems*. New Ed edition. Cambridge, Mass.: MIT Press.

Eliasmith, C. & Kolbeck, C. (2015) Marr's Attacks: On Reductionism and Vagueness. *Topics in Cognitive Science*. 7 (2), 323–335.

Elman, J.L. (1993) Learning and development in neural networks: the importance of starting small. *Cognition*. 48 (1), 71–99.

Endress, A.D., Carden, S., Versace, E. & Hauser, M.D. (2010) The apes' edge: positional learning in chimpanzees and humans. *Animal Cognition*. 13 (3), 483–495.

Endress, A.D., Nespor, M. & Mehler, J. (2009) Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*. 13 (8), 348–353.

Erez, J., Cusack, R., Kendall, W. & Barense, M.D. (2016) Conjunctive Coding of Complex Object Features. *Cerebral Cortex (New York, NY)*. 26 (5), 2271–2282.

Esser, S. & Haider, H. (2017) The Emergence of Explicit Knowledge in a Serial Reaction Time Task: The Role of Experienced Fluency and Strength of Representation. *Frontiers in Psychology*. 8, 502.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M. & Wong, P.C.M. (2016) The Relationship between Artificial and Second Language Learning. *Cognitive science*. 40 (4), 822–847.

Evans, J.L., Saffran, J.R. & Robe-Torres, K. (2009) Statistical Learning in Children With Specific Language Impairment. *Journal of speech, language, and hearing research : JSLHR*. 52 (2), 10.1044/1092-4388(2009/07–0189).

Faraco, C.C., Unsworth, N., Langley, J., Terry, D., Li, K., Zhang, D., Liu, T. & Miller, L.S. (2011) Complex span tasks and hippocampal recruitment during working memory. *NeuroImage*. 55 (2), 773–787.

Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*. 39 (2), 175–191.

Feldman, J. (2013) The neural binding problem(s). *Cognitive Neurodynamics*. 7 (1), 1–11.

Fiser, J. & Aslin, R.N. (2002) Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. 28 (3), 458–467.

Fitch, W.T. (2014) Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*. 11 (3), 329–364.

Fitch, W.T. & Friederici, A.D. (2012) Artificial grammar learning meets formal language theory: an overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 367 (1598), 1933–1955.

Fitch, W.T. & Martins, M.D. (2014) Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*. 1316 (1), 87–104.

Fló, A. (2021) Evidence of ordinal position encoding of sequences extracted from continuous speech. *Cognition*. 213, 104646.

Folia, V. & Petersson, K.M. (2014) Implicit structured sequence learning: an fMRI study of the structural mere-exposure effect. *Cognition*. 5, 41.

Fonov, V., Evans, A., McKinstry, R., Almli, C. & Collins, D. (2009) Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*. 47, S102.

Forkstam, C., Elwér, Å., Ingvar, M. & Petersson, K.M. (2008) Instruction effects in implicit artificial grammar learning: A preference for grammaticality. *Brain Research*. 1221, 80–92.

Fox, N.P., Leonard, M., Sjerps, M.J. & Chang, E.F. (2020) Transformation of a temporal speech cue to a spatial neural code in human auditory cortex Jonathan Erik Peelle, Barbara G Shinn-Cunningham, & Michael Wehr (eds.). *eLife*. 9, e53051.

Friederici, A.D. (2020) Hierarchy processing in human neurobiology: how specific is it? *Philosophical Transactions of the Royal Society B: Biological Sciences*. 375 (1789), 20180391.

Friederici, A.D. (2004) Processing local transitions versus long-distance syntactic hierarchies. *Trends in Cognitive Sciences*. 8 (6), 245–247.

Friederici, A.D. (2011) The brain basis of language processing: from structure to function. *Physiological Reviews*. 91 (4), 1357–1392.

Friederici, A.D. (2012) The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences*. 16 (5), 262–268.

Friederici, A.D. (2002) Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*. 6 (2), 78–84.

Friederici, A.D., Bahlmann, J., Heim, S., Schubotz, R.I. & Anwander, A. (2006) The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*. 103 (7), 2458–2463.

Friederici, A.D. & Chomsky, N. (2017) *Language in Our Brain: The Origins of a Uniquely Human Capacity*. 1 edition. Cambridge, Massachusetts: The MIT Press.

Friederici, A.D., von Cramon, D.Y. & Kotz, S.A. (1999) Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain*. 122 (6), 1033–1047.

Friederici, A.D., Fiebach, C.J., Schlesewsky, M., Bornkessel, I.D. & von Cramon, D.Y. (2006) Processing Linguistic Complexity and Grammaticality in the Left Frontal Cortex. *Cerebral Cortex*. 16 (12), 1709–1717.

Friederici, A.D. & Gierhan, S.M.E. (2013) The language network. *Current Opinion in Neurobiology*. 23 (2), 250–254.

Friederici, A.D., Makuuchi, M. & Bahlmann, J. (2009) The role of the posterior superior temporal cortex in sentence comprehension. *NeuroReport*. 20 (6), 563–568.

Friederici, A.D. & Meyer, M. (2004) The brain knows the difference: two types of grammatical violations. *Brain Research*. 1000 (1), 72–77.

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

Friederici, A.D., Pfeifer, E. & Hahne, A. (1993) Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*. 1 (3), 183–192.

Friederici, A.D., Steinhauer, K. & Pfeifer, E. (2002) Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences*. 99 (1), 529–534.

Friedrich, R. & Friederici, A.D. (2009) Mathematical Logic in the Human Brain: Syntax. *PLOS ONE*. 4 (5), e5599.

Fries, P. (2015) Rhythms for Cognition: Communication through Coherence. *Neuron*. 88 (1), 220–235.

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

Frost, R., Armstrong, B.C., Siegelman, N. & Christiansen, M.H. (2015) Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*. 19 (3), 117–125.

Fuji, M., Maesawa, S., Ishiai, S., Iwami, K., Futamura, M. & Saito, K. (2016) Neural Basis of Language: An Overview of An Evolving Model. *Neurologia medico-chirurgica*. 56 (7), 379–386.

Galluppi, F., Conradt, J., Stewart, T., Eliasmith, C., Horiuchi, T., Tapson, J., Tripp, B., Furber, S. & Etienne-Cummings, R. (2012) 'Live Demo: Spiking ratSLAM: Rat hippocampus cells in spiking neural hardware', in *2012 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. [Online]. November 2012. pp. 91–91.

Gansner, E.R., Koutsofios, E., North, S.C. & Vo, K.-P. (1993) A technique for drawing directed graphs. *IEEE Transactions on Software Engineering*. 19 (3), 214–230.

Gao, R., van den Brink, R.L., Pfeffer, T. & Voytek, B. (2020) Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. Martin Vinck, Laura L Colgin, & Thilo Womelsdorf (eds.). *eLife*. 9, e61277.

Gasser, B. & Arbib, M.A. (2017) A neuro-computational model of sequence learning in macaques: the Simultaneous Chaining Paradigm. *Adaptive Behavior*. 25 (4), 195–213.

Gayler, R.W. (2003) 'Vector Symbolic Architectures answer Jackendoff's challenges for cognitive neuroscience', in *Proceedings of the ICCS/ASCS International Conference on Cognitive Science*. pp. 133–138.

Gebhart, A.L., Newport, E.L. & Aslin, R.N. (2009) Statistical Learning of Adjacent and Non-Adjacent Dependencies among Non-Linguistic Sounds. *Psychonomic bulletin & review*. 16 (3), 486–490.

Gervain, J. & Werker, J.F. (2013) Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*. 4 (1), 1490.

Geweke, J. (1982) Measurement of Linear Dependence and Feedback between Multiple Time Series. *Journal of the American Statistical Association*. 77 (378), 304–313.

Geweke, J.F. (1984) Measures of Conditional Linear Dependence and Feedback between Time Series. *Journal of the American Statistical Association*. 79 (388), 907–915.

Gheysen, F., Van Opstal, F., Roggeman, C., Van Waelvelde, H. & Fias, W. (2010) Hippocampal contribution to early and later stages of implicit motor sequence learning. *Experimental Brain Research*. 202 (4), 795–807.

Ghirlanda, S., Lind, J. & Enquist, M. (2017) Memory for stimulus sequences: a divide between humans and other animals? *Royal Society Open Science*. 4 (6), 161011.

Ghitza, O. (2011) Linking Speech Perception and Neurophysiology: Speech Decoding Guided by Cascaded Oscillators Locked to the Input Rhythm. *Frontiers in Psychology*. 2, 130.

Gilra, A. & Gerstner, W. (2017) Predicting non-linear dynamics by stable local learning in a recurrent spiking neural network. *eLife*. 6, e28295.

Giorgi, F.S., Maestri, M., Guida, M., Di Coscio, E., Carnicelli, L., Perini, D., Pizzanelli, C., Iudice, A. & Bonanni, E. (2013) Controversial Issues on EEG after Sleep Deprivation for the Diagnosis of Epilepsy. *Epilepsy Research and Treatment*. 2013, 614685.

Giraud, A.-L. & Poeppel, D. (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*. 15 (4), 511–517.

Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S. & Ellis, R. (2015) Timed and Untimed Grammaticality Judgments Measure Distinct Types of Knowledge: Evidence from Eye-Movement Patterns. *Studies in Second Language Acquisition*. 37 (2), 269–297.

Gómez, R.L. (2002) Variability and detection of invariant structure. *Psychological Science*. 13 (5), 431–436.

Gonda, S., Tarrasch, R. & Ben Shalom, D. (2020) The functional significance of the P600. *Medicine*. 99 (46), e23116.

Goschke, T., Friederici, A.D., Kotz, S.A. & van Kampen, A. (2001) Procedural learning in Broca's aphasia: dissociation between the implicit acquisition of spatio-motor and phoneme sequences. *Journal of Cognitive Neuroscience*. 13 (3), 370–388.

Gosmann, J. & Eliasmith, C. (2019) Vector-Derived Transformation Binding: An Improved Binding Operation for Deep Symbol-Like Processing in Neural Networks. *Neural Computation*. 31 (5), 849–869.

Gouvea, A.C., Phillips, C., Kazanina, N. & Poeppel, D. (2010) The linguistic processes underlying the P600. *Language and Cognitive Processes*. 25 (2), 149–188.

Gower, J.C. (1975) Generalized procrustes analysis. *Psychometrika*. 40 (1), 33–51.

Gower, J.C. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*. 53 (3–4), 325–338.

Grama, I.C., Kerkhoff, A. & Wijnen, F. (2016) Gleaning Structure from Sound: The Role of Prosodic Contrast in Learning Non-adjacent Dependencies. *Journal of Psycholinguistic Research*. 45 (6), 1427–1449.

Grandchamp, R. & Delorme, A. (2011) Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials. *Frontiers in Psychology*. 2, 236.

Granger, C.W.J. (1969) Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*. 37 (3), 424–438.

Gray, C.M. (1999) The Temporal Correlation Hypothesis of Visual Feature Integration: Still Alive and Well. *Neuron*. 24 (1), 31–47.

Grech, R., Cassar, T., Muscat, J., Camilleri, K.P., Fabri, S.G., Zervakis, M., Xanthopoulos, P., Sakkalis, V. & Vanrumste, B. (2008) Review on solving the inverse problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation*. 5 (1), 25.

Griffiths, B.J., Parish, G., Roux, F., Michelmann, S., Plas, M. van der, Kolibius, L.D., Chelvarajah, R., Rollings, D.T., Sawlani, V., Hamer, H., Gollwitzer, S., Kreiselmeyer, G., Staresina, B., Wimber, M. & Hanslmayr, S. (2019) Directional coupling of slow and fast hippocampal gamma with neocortical alpha/beta oscillations in human episodic memory. *Proceedings of the National Academy of Sciences*. 116 (43), 21834–21842.

Grodzinsky, Y. (2000) The neurology of syntax: language use without Broca's area. *The Behavioral and Brain Sciences*. 23 (1), 1–21; discussion 21-71.

Gunter, T.C., Friederici, A.D. & Schriefers, H. (2000) Syntactic Gender and Semantic Expectancy: ERPs Reveal Early Autonomy and Late Interaction. *Journal of Cognitive Neuroscience*. 12 (4), 556–568.

Gunter, T.C., Stowe, L.A. & Mulder, G. (1997) When syntax meets semantics. *Psychophysiology*. 34 (6), 660–676.

Haberlandt, K., Thomas, J.G., Lawrence, H. & Krohn, T. (2005) Transposition asymmetry in immediate serial recall. *Memory*. 13 (3–4), 274–282.

Hadley, R.F. (2007) Synchronous versus conjunctive binding: a false dichotomy? *Connection Science*. 19 (2), 111–130.

Hagoort, P. (2013) MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*. 4, 416.

Hagoort, P. (2005) On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*. 9 (9), 416–423.

Hagoort, P., Brown, C. & Groothusen, J. (1993) The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*. 8 (4), 439–483.

Halford, G.S., Wilson, W.H. & Phillips, S. (1998) Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *The Behavioral and Brain Sciences*. 21 (6), 803–831; discussion 831-864.

Hall, M.L., Eigsti, I.-M., Bortfeld, H. & Lillo-Martin, D. (2018) Auditory access, language access, and implicit sequence learning in deaf children. *Developmental science*. 21 (3), e12575.

Hamilton, L.S. & Huth, A.G. (2020) The revolution will not be controlled: natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*. 35 (5), 573–582.

Hansen, N.C. & Pearce, M.T. (2014) Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*. 5, 1052.

Harris, Z.S. (1957) Co-occurrence and Transformation in Linguistic Structure. *Language*. 33 (3), 283–340.

Hasselmo, M.E., Alexander, A.S., Dannenberg, H. & Newman, E.L. (2020) Overview of computational models of hippocampus and related structures: Introduction to the special issue. *Hippocampus*. 30 (4), 295–301.

Haufe, S., DeGuzman, P., Henin, S., Arcaro, M., Honey, C.J., Hasson, U. & Parra, L.C. (2018) Elucidating relations between fMRI, ECoG, and EEG through a common natural stimulus. *NeuroImage*. 179, 79–91.

Hauk, O. & Weiss, B. (2020) The neuroscience of natural language processing. *Language, Cognition and Neuroscience*. 35 (5), 541–542.

Hauser, M.D., Chomsky, N. & Fitch, W.T. (2002) The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*. 298 (5598), 1569–1579.

Hawthorne, K. & Gerken, L. (2014) From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*. 133 (2), 420–428.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L. & Pietrini, P. (2001) Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*. 293 (5539), 2425–2430.

Haynes, J.-D. & Rees, G. (2006) Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*. 7 (7), 523–534.

Heimbauer, L.A., Conway, C.M., Christiansen, M.H., Beran, M.J. & Owren, M.J. (2012) A Serial Reaction Time (SRT) task with symmetrical joystick responding for nonhuman primates. *Behavior Research Methods*. 44 (3), 733–741.

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

Henson, R.N.A. (1999) Positional information in short-term memory: Relative or absolute? *Memory & Cognition*. 27 (5), 915–927.

Henson, R.N.A. & Burgess, N. (1998) 'Representations of Serial Order', in John A. Bullinaria, David W. Glasspool, & George Houghton (eds.) *4th Neural Computation and Psychology*

Workshop, London, 9–11 April 1997. Perspectives in Neural Computing. [Online]. 1998 London: Springer. pp. 283–300.

Herbranson, W.T. & Shimp, C.P. (2008) Artificial grammar learning in pigeons. *Learning & Behavior*. 36 (2), 116–137.

Herrmann, C.S. & Knight, R.T. (2001) Mechanisms of human attention: event-related potentials and oscillations. *Neuroscience & Biobehavioral Reviews*. 25 (6), 465–476.

Herrmann, C.S., Rach, S., Vosskuhl, J. & Strüber, D. (2014) Time–Frequency Analysis of Event-Related Potentials: A Brief Tutorial. *Brain Topography*. 27 (4), 438–450.

Hertrich, I., Dietrich, S. & Ackermann, H. (2020) The Margins of the Language Network in the Brain. *Frontiers in Communication*. 5, 519955.

Himmelstoss, N.A., Schuster, S., Hutzler, F., Moran, R. & Hawelka, S. (2020) Co-registration of eye movements and neuroimaging for studying contextual predictions in natural reading. *Language, Cognition and Neuroscience*. 35 (5), 595–612.

Hochreiter, S. & Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Comput.* 9 (8), 1735–1780.

Hodgkin, A.L. & Huxley, A.F. (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*. 117 (4), 500–544.

Hoen, M. & Dominey, P.F. (2000) ERP analysis of cognitive sequencing: a left anterior negativity related to structural transformation processing. *NeuroReport*. 11 (14), 3187–3191.

Hogendoorn, H. & Burkitt, A.N. (2019) Predictive Coding with Neural Transmission Delays: A Real-Time Temporal Alignment Hypothesis. *eNeuro*. 6 (2), ENEURO.0412-18.2019.

Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*. 24 (6), 417–441.

Hotelling, H. (1953) New Light on the Correlation Coefficient and its Transforms. *Journal of the Royal Statistical Society. Series B (Methodological)*. 15 (2), 193–232.

Howard, M.W. & Eichenbaum, H. (2013) The hippocampus, time, and memory across scales. *Journal of Experimental Psychology: General*. 142 (4), 1211–1230.

Hsu, H.J., Tomblin, J.B. & Christiansen, M.H. (2014) Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Frontiers in Psychology*. 5, 175.

Hsu, N.S., Jaeggi, S.M. & Novick, J.M. (2017) A common neural hub resolves syntactic and non-syntactic conflict through cooperation with task-specific networks. *Brain and language*. 166, 63–77.

Hummel, J.E. & Holyoak, K.J. (2003) A symbolic-connectionist theory of relational inference and generalization. *Psychological review*. 110 (2), 220–264.

Hummel, J.E. & Holyoak, K.J. (1997) Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*. 104 (3), 427–466.

Hummel, J.E., Holyoak, K.J., Green, C., Doumas, L.A.A., Devnich, D., Aniket, Kittur, A. & Kalar, D.J. (2004) 'A solution to the binding problem for compositional connectionism', in *AAAI 2004*. [Online]. 2004 AAAI Press. pp. 31–34.

Hurvich, C.M. & Tsai, C. (1989) Regression and time series model selection in small samples. *Biometrika*. 76 (2), 297–307.

Izhikevich, E.M. (2003) Simple model of spiking neurons. *IEEE transactions on neural networks*. 14 (6), 1569–1572.

Jablonowski, J., Taesler, P., Fu, Q. & Rose, M. (2018) Implicit acoustic sequence learning recruits the hippocampus. *PLOS ONE*. 13 (12), e0209590.

Jackendoff, R. & Pinker, S. (2005) The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*. 97 (2), 211–225.

Jeon, H.-A. (2014) Hierarchical processing in the prefrontal cortex in a variety of cognitive domains. *Frontiers in Systems Neuroscience*. 8, 223.

Jiang, X., Long, T., Cao, W., Li, J., Dehaene, S. & Wang, L. (2018) Production of Supra-regular Spatial Sequences by Macaque Monkeys. *Current Biology*. 28 (12), 1851-1859.e4.

Johnson, E.K. & Jusczyk, P.W. (2001) Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language*. 44 (4), 548–567.

Kachergis, G., Yu, C. & Shiffrin, R.M. (2014) Cross-situational word learning is both implicit and strategic. *Frontiers in Psychology*. 5, 588.

Kaestner, E., Morgan, A.M., Snider, J., Zhan, M., Jiang, X., Levy, R., Ferreira, V.S., Thesen, T. & Halgren, E. (2020) Toward A database of intracranial electrophysiology during natural language presentation. *Language, Cognition and Neuroscience*. 35 (6), 729–738.

Kahta, S. & Schiff, R. (2016) Implicit learning deficits among adults with developmental dyslexia. *Annals of Dyslexia*. 66 (2), 235–250.

Kanerva, P. (1994) 'The Spatter Code for Encoding Concepts at Many Levels', in Maria Marinaro & Pietro G. Morasso (eds.) *ICANN '94*. [Online]. 1994 Springer London. pp. 226–229.

Kar, K., Kubilius, J., Schmidt, K., Issa, E.B. & DiCarlo, J.J. (2019) Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience*. 22 (6), 974–983.

Kaur, K., Shih, J. & Krusienski, D. (2013) 'Modeling the Electroencephalogram Using Intracranial Signals', in *Proceedings of the Fifth International Brain-Computer Interface Meeting: Defining the Future*. [Online]. 2013 Pacific Grove, California. Article 103.

Kemény, F. & Németh, K. (2017) Stimulus dependence and cross-modal interference in sequence learning. *The Quarterly Journal of Experimental Psychology*. 70 (12), 2535–2547.

Kemler Nelson, D.G., Jusczyk, P.W., Mandel, D.R., Myers, J., Turk, A. & Gerken, L. (1995) The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*. 18 (1), 111–116.

Kepinska, O., de Rover, M., Caspers, J. & Schiller, N.O. (2018) Connectivity of the hippocampus and Broca's area during acquisition of a novel grammar. *NeuroImage*. 165, 1–10.

ter Keurs, M., Brown, C.M., Hagoort, P. & Stegeman, D.F. (1999) Electrophysiological manifestations of open- and closed-class words in patients with Broca's aphasia with agrammatic comprehension: An event-related brain potential study. *Brain*. 122 (5), 839–854.

Kielar, A., Deschamps, T., Jokel, R. & Meltzer, J.A. (2018) Abnormal language-related oscillatory responses in primary progressive aphasia. *NeuroImage: Clinical*. 18, 560–574.

Kielar, A., Meltzer, J.A., Moreno, S., Alain, C. & Bialystok, E. (2014) Oscillatory Responses to Semantic and Syntactic Violations. *Journal of Cognitive Neuroscience*. 26 (12), 2840–2862.

Kielar, A., Panamsky, L., Links, K.A. & Meltzer, J.A. (2015) Localization of electrophysiological responses to semantic and syntactic anomalies in language comprehension with MEG. *NeuroImage*. 105, 507–524.

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

Kikuchi, Y., Ip, J., Lagier, G., Mossom, J.C., Kumar, S., Petkov, C.I., Barraclough, N.E. & Vuong, Q.C. (2019) Interactions between conscious and subconscious signals: Selective attention under feature-based competition increases neural selectivity during brain adaptation. *Journal of Neuroscience*. 39 (28), 5506–5516.

Kikuchi, Y., Sedley, W., Griffiths, T.D. & Petkov, C.I. (2018) Evolutionarily conserved neural signatures involved in sequencing predictions and their relevance for language. *Current Opinion in Behavioral Sciences*. 21, 145–153.

Klein, K.A., Addis, K.M. & Kahana, M.J. (2005) A comparative analysis of serial and free recall. *Memory & Cognition*. 33 (5), 833–839.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R. & Broussard, C. (2007) What's new in psychtoolbox-3. *Perception*. 36 (14), 1–16.

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., Ringelstein, E.B. & Henningsen, H. (2000) Handedness and hemispheric language dominance in healthy humans. *Brain: A Journal of Neurology*. 123 (12), 2512–2518.

Koch, F.-S., Sundqvist, A., Thornberg, U.B., Ullman, M.T., Barr, R., Rudner, M. & Heimann, M. (2020) Data and analysis script for infant and adult eye movement in an adapted ocular-motor serial reaction time task assessing procedural memory. *Data in Brief*. 29, 105108.

Koechlin, E. & Jubault, T. (2006) Broca's Area and the Hierarchical Organization of Human Behavior. *Neuron*. 50 (6), 963–974.

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. *Scientific Reports*. 6 (1), 19741.

Koelsch, S., Rohrmeier, M., Torrecuso, R. & Jentschke, S. (2013) Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences*. 110 (38), 15443–15448.

Koelsch, S. & Siebel, W.A. (2005) Towards a neural basis of music perception. *Trends in Cognitive Sciences*. 9 (12), 578–584.

Komorowski, R.W., Manns, J.R. & Eichenbaum, H. (2009) Robust Conjunctive Item–Place Coding by Hippocampal Neurons Parallels Learning What Happens Where. *Journal of Neuroscience*. 29 (31), 9918–9929.

Kong, J., Wang, C., Kwong, K., Vangel, M., Chua, E. & Gollub, R. (2005) The neural substrate of arithmetic operations and procedure complexity. *Cognitive Brain Research*. 22 (3), 397–405.

Kovach, C.K. & Gander, P.E. (2016) The demodulated band transform. *Journal of Neuroscience Methods*. 261, 135–154.

Kraus, B.J., Robinson, R.J., White, J.A., Eichenbaum, H. & Hasselmo, M.E. (2013) Hippocampal 'time cells': time versus path integration. *Neuron*. 78 (6), 1090–1101.

Kriegeskorte, N. & Douglas, P.K. (2019) Interpreting encoding and decoding models. *Current opinion in neurobiology*. 55, 167–179.

Kriegeskorte, N., Mur, M. & Bandettini, P. (2008) Representational Similarity Analysis – Connecting the Branches of Systems Neuroscience. *Frontiers in Systems Neuroscience*. 2, 4.

Krusienski, D.J. & Shih, J.J. (2010) 'A case study on the relation between electroencephalographic and electrocorticographic event-related potentials', in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. [Online]. August 2010. pp. 6019–6022.

Kubanek, J. & Schalk, G. (2015) NeuralAct: A Tool to Visualize Electrocortical (ECoG) Activity on a Three-Dimensional Model of the Cortex. *Neuroinformatics*. 13 (2), 167–174.

Kuchibhotla, K. & Bathellier, B. (2018) Neural encoding of sensory and behavioral complexity in the auditory cortex. *Current Opinion in Neurobiology*. 52, 65–71.

Kumar, S., Gander, P.E., Berger, J.I., Billig, A.J., Nourski, K.V., Oya, H., Kawasaki, H., Howard, M.A. & Griffiths, T.D. (2020) Oscillatory correlates of auditory working memory examined with human electrocorticography. *Neuropsychologia*. 150, 107691.

Kumar, S., Joseph, S., Gander, P.E., Barascud, N., Halpern, A.R. & Griffiths, T.D. (2016) A Brain System for Auditory Working Memory. *The Journal of Neuroscience*. 36 (16), 4492–4505.

Kuntzelman, K.M., Williams, J.M., Lim, P.C., Samal, A., Rao, P.K. & Johnson, M.R. (2021) Deep-Learning-Based Multivariate Pattern Analysis (dMVPA): A Tutorial and a Toolbox. *Frontiers in Human Neuroscience*. 15, 638052.

Kuppuraj, S., Duta, M., Thompson, P. & Bishop, D. (2018) Online incidental statistical learning of audiovisual word sequences in adults: a registered report. *Royal Society Open Science*. 5 (2), 171678.

Kutas, M. & Federmeier, K.D. (2011) Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual review of psychology*. 62, 621–647.

Kutas, M. & Hillyard, S.A. (1989) An Electrophysiological Probe of Incidental Semantic Association. *Journal of Cognitive Neuroscience*. 1 (1), 38–49.

Kutas, M. & Hillyard, S.A. (1983) Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*. 11 (5), 539–550.

Kutas, M. & Hillyard, S.A. (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*. 207 (4427), 203–205.

LaCross, A. (2015) Khalkha Mongolian speakers' vowel bias: L1 influences on the acquisition of non-adjacent vocalic dependencies. *Language, Cognition and Neuroscience*. 30 (9), 1033–1047.

Lai, J. & Poletiek, F.H. (2011) The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition*. 118 (2), 265–273.

Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I. & Schroeder, C.E. (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science (New York, N.Y.)*. 320 (5872), 110–113.

Lande, R.G. & Gragnani, C. (2015) Relationships Between Polypharmacy and the Sleep Cycle Among Active-Duty Service Members. *The Journal of the American Osteopathic Association*. 115 (6), 370–375.

Lany, J. & Gómez, R.L. (2008) Twelve-Month-Old Infants Benefit From Prior Experience in Statistical Learning. *Psychological Science*. 19 (12), 1247–1252.

Lany, J., Gómez, R.L. & Gerken, L.A. (2007) The Role of Prior Experience in Language Acquisition. *Cognitive Science*. 31 (3), 481–507.

Lanzi, O. (2016) *Metric for how symmetric a matrix is*. [Online]. Available from: https://math.stackexchange.com/q/2048830.

Lashley, K.S. (1951) 'The problem of serial order in behavior', in *Cerebral mechanisms in behavior; the Hixon Symposium*. [Online]. Oxford, England: Wiley. pp. 112–146.

Lelekov-Boissard, T. & Dominey, P.F. (2002) Human brain potentials reveal similar processing of non-linguistic abstract structure and linguistic syntactic structure. *Neurophysiologie Clinique/Clinical Neurophysiology*. 32 (1), 72–84.

Levy, D.F. & Wilson, S.M. (2020) Categorical Encoding of Vowels in Primary Auditory Cortex. *Cerebral Cortex (New York, N.Y.: 1991)*. 30 (2), 618–627.

Levy, S.D. & Gayler, R. (2008) 'Vector Symbolic Architectures: A New Building Material for Artificial General Intelligence', in *Proceedings of the 2008 Conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference*. [Online]. Amsterdam, The Netherlands, The Netherlands: IOS Press. pp. 414–418.

Lewis, A.G. & Bastiaansen, M. (2015) A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*. 68, 155–168.

Lewis, A.G., Lemhöfer, K., Schoffelen, J.-M. & Schriefers, H. (2016) Gender agreement violations modulate beta oscillatory dynamics during sentence comprehension: A comparison of second language learners and native speakers. *Neuropsychologia*. 89, 254–272.

Lewis, A.G., Wang, L. & Bastiaansen, M. (2015) Fast oscillatory dynamics during language comprehension: Unification versus maintenance and prediction? *Brain and Language*. 148, 51–63.

Li, X., Zhao, X., Shi, W., Lu, Y. & Conway, C.M. (2018) Lack of Cross-Modal Effects in Dual-Modality Implicit Statistical Learning. *Frontiers in Psychology*. 9, 146.

Liégeois-Chauvel, C., Lorenzi, C., Trébuchon, A., Régis, J. & Chauvel, P. (2004) Temporal envelope processing in the human left and right auditory cortices. *Cerebral Cortex (New York, N.Y.: 1991)*. 14 (7), 731–740.

Lin, B., Mur, M., Kietzmann, T. & Kriegeskorte, N. (2019) 'Visualizing representational dynamics with multidimensional scaling alignment', in *Proceedings of the 2019 Conference on Cognitive Computational Neuroscience*. [Online]. pp. 1030–1033.

Lisman, J.E. & Jensen, O. (2013) The Theta-Gamma Neural Code. *Neuron*. 77 (6), 1002–1016.

Liu, C. (2013) Just noticeable difference of tone pitch contour change for English- and Chinese-native listeners. *The Journal of the Acoustical Society of America*. 134 (4), 3011–3020.

Logie, R. (2007) Google-Books-ID: I4CtrmoyeOQC. *The Cognitive Neuroscience of Working Memory*. Oxford University Press.

Long, N.M. & Kahana, M.J. (2019) Hippocampal contributions to serial-order memory. *Hippocampus*. 29 (3), 252–259.

Loui, P., Li, H.C. & Schlaug, G. (2011) White matter integrity in right hemisphere predicts pitch-related grammar learning. *NeuroImage*. 55 (2), 500–507.

Lu, T., Voelker, A.R., Komer, B. & Eliasmith, C. (2019) 'Representing spatial relations with fractional binding', in *Proceedings of the Cognitive Science Society*. [Online]. pp. 2214–2220.

Luck, S.J. & Kappenman, E.S. (2013) Google-Books-ID: p_sTDAAAQBAJ. *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press.

Lundqvist, M., Herman, P. & Miller, E.K. (2018) Working Memory: Delay Activity, Yes! Persistent Activity? Maybe Not. *Journal of Neuroscience*. 38 (32), 7013–7019.

Luo, H. & Poeppel, D. (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*. 54 (6), 1001–1010.

Luo, Y., Zhang, Y., Feng, X. & Zhou, X. (2010) Electroencephalogram oscillations differentiate semantic and prosodic processes during sentence reading. *Neuroscience*. 169 (2), 654–664.

Makeig, S. (1993) Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and Clinical Neurophysiology*. 86 (4), 283–293.

Makeig, S., Debener, S., Onton, J. & Delorme, A. (2004) Mining event-related brain dynamics. *Trends in Cognitive Sciences*. 8 (5), 204–210.

Manting, C.L., Gulyas, B., Ullén, F. & Lundqvist, D. (2021) Auditory steady-state responses during and after a stimulus: Cortical sources, and the influence of attention and musicality. *NeuroImage*. 233, 117962.

Marblestone, A.H., Wayne, G. & Kording, K.P. (2016) Toward an Integration of Deep Learning and Neuroscience. *Frontiers in Computational Neuroscience*. 10, 94.

Marcus, G.F., Vijayan, S., Rao, S.B. & Vishton, P.M. (1999) Rule Learning by Seven-Month-Old Infants. *Science*. 283 (5398), 77–80.

Maris, E. & Oostenveld, R. (2007) Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*. 164 (1), 177–190.

Marr, D. (1982) *Vision: A computational approach*. Freeman. San Francisco.

Marti, S., Thibault, L. & Dehaene, S. (2014) How Does the Extraction of Local and Global Auditory Regularities Vary with Context? *PLOS ONE*. 9 (9), e107227.

Martin, A.E. & Doumas, L.A.A. (2017) A mechanism for the cortical computation of hierarchical linguistic structure. *PLOS Biology*. 15 (3), e2000663.

Martin, S., Iturrate, I., Millán, J. del R., Knight, R.T. & Pasley, B.N. (2018) Decoding Inner Speech Using Electrocorticography: Progress and Challenges Toward a Speech Prosthesis. *Frontiers in Neuroscience*. 12, 422.

Matchin, W., Hammerly, C. & Lau, E. (2017) The role of the IFG and pSTS in syntactic prediction: Evidence from a parametric study of hierarchical structure in fMRI. *Cortex*. 88, 106–123.

Matchin, W. & Hickok, G. (2020) The Cortical Organization of Syntax. *Cerebral Cortex (New York, N.Y.: 1991)*. 30 (3), 1481–1498.

van de Meerendonk, N., Indefrey, P., Chwilla, D.J. & Kolk, H.H.J. (2011) Monitoring in language perception: Electrophysiological and hemodynamic responses to spelling violations. *NeuroImage*. 54 (3), 2350–2363.

Miller, G.A. (1958) Free recall of redundant strings of letters. *Journal of Experimental Psychology*. 56 (6), 485–491.

Milne, A.E., Petkov, C.I. & Wilson, B. (2018) Auditory and Visual Sequence Learning in Humans and Monkeys using an Artificial Grammar Learning Paradigm. *Neuroscience*. 389, 104–117.

Misyak, J.B. & Christiansen, M.H. (2012) Statistical Learning and Language: An Individual Differences Study. *Language Learning*. 62 (1), 302–331.

Misyak, J.B., Christiansen, M.H. & Bruce Tomblin, J. (2010) Sequential Expectations: The Role of Prediction-Based Learning in Language. *Topics in Cognitive Science*. 2 (1), 138–153.

Moga, S. & Gaussier, P. (2003) 'Artificial Neural Network for Sequence Learning', in *IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003*. [Online]. p. 1507.

Molenberghs, P., Cunnington, R. & Mattingley, J.B. (2012) Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies. *Neuroscience & Biobehavioral Reviews*. 36 (1), 341–349.

Molinaro, N., Barber, H.A. & Carreiras, M. (2011) Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*. 47 (8), 908–930.

Molinaro, N., Vespignani, F., Zamparelli, R. & Job, R. (2011) Why brother and sister are not just siblings: Repair processes in agreement computation. *Journal of Memory and Language*. 64 (3), 211–232.

Monaghan, P., Christiansen, M.H. & Chater, N. (2007) The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*. 55 (4), 259–305.

Morgan, E.U., van der Meer, A., Vulchanova, M., Blasi, D.E. & Baggio, G. (2020) Meaning before grammar: A review of ERP experiments on the neurodevelopmental origins of semantic processing. *Psychonomic Bulletin & Review*. 27 (3), 441–464.

Moser, M.-B., Rowland, D.C. & Moser, E.I. (2015) Place Cells, Grid Cells, and Memory. *Cold Spring Harbor Perspectives in Biology*. 7 (2), a021808.

Mueller, J.L., Friederici, A.D. & Männel, C. (2012) Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences of the United States of America*. 109 (39), 15953–15958.

Münte, T.F., Heinze, H.-J. & Mangun, G.R. (1993) Dissociation of Brain Activity Related to Syntactic and Semantic Aspects of Language. *Journal of Cognitive Neuroscience*. 5 (3), 335–344.

Murdock, B.B. (1962) The serial position effect of free recall. *Journal of Experimental Psychology*. 64 (5), 482–488.

Nelson, M.J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., Cash, S.S., Naccache, L., Hale, J.T., Pallier, C. & Dehaene, S. (2017) Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences of the United States of America*. 114 (18), E3669–E3678.

Neville, H., Nicol, J.L., Barss, A., Forster, K.I. & Garrett, M.F. (1991) Syntactically Based Sentence Processing Classes: Evidence from Event-Related Brain Potentials. *Journal of Cognitive Neuroscience*. 3 (2), 151–165.

Newport, E.L. & Aslin, R.N. (2004) Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*. 48 (2), 127–162.

Nieder, A. (2012) Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proceedings of the National Academy of Sciences*. 109 (29), 11860–11865.

Nieh, E.H., Schottdorf, M., Freeman, N.W., Low, R.J., Lewallen, S., Koay, S.A., Pinto, L., Gauthier, J.L., Brody, C.D. & Tank, D.W. (2021) Geometry of abstract learned knowledge in the hippocampus. *Nature*. 595 (7865), 80–84.

Ninokura, Y., Mushiake, H. & Tanji, J. (2004) Integration of Temporal Order and Object Information in the Monkey Lateral Prefrontal Cortex. *Journal of Neurophysiology*. 91 (1), 555–560.

Nissen, M.J. & Bullemer, P. (1987) Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*. 19 (1), 1–32.

Nordby, H., Hugdahl, K., Stickgold, R., Bronnick, K.S. & Hobson, J.A. (1996) Event-related potentials (ERPs) to deviant auditory stimuli during sleep and waking. *NeuroReport*. 7 (5), 1082–1086.

Norman, K.A., Polyn, S.M., Detre, G.J. & Haxby, J.V. (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*. 10 (9), 424–430.

Nourski, K.V. & Howard, M.A. (2015) Invasive recordings in the human auditory cortex. *Handbook of Clinical Neurology*. 129, 225–244.

Nourski, K.V., Reale, R.A., Oya, H., Kawasaki, H., Kovach, C.K., Chen, H., Howard, M.A. & Brugge, J.F. (2009) Temporal envelope of time-compressed speech represented in the human auditory cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 29 (49), 15564–15574.

Nourski, K.V., Steinschneider, M., Rhone, A.E., Kovach, C.K., Banks, M.I., Krause, B.M., Kawasaki, H. & Howard, M.A., III (2021) Electrophysiology of the Human Superior Temporal Sulcus during Speech Processing. *Cerebral Cortex*. 31 (2), 1131–1148.

Oganian, Y. & Chang, E.F. (2019) A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances*. 5 (11), eaay6279.

O'Keefe, J. (1976) Place units in the hippocampus of the freely moving rat. *Experimental Neurology*. 51 (1), 78–109.

Onnis, L., Monaghan, P., Christiansen, M.H. & Chater, N. (2004) 'Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies', in *Proceedings of the Annual Meeting of the Cognitive Science Society*. [Online]. pp. 1047–1052.

Onnis, L., Monaghan, P., Richmond, K. & Chater, N. (2005) Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*. 53 (2), 225–237.

Onnis, L. & Thiessen, E. (2013) Language experience changes subsequent learning. *Cognition*. 126 (2), 268–284.

Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. (2011) FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*. 2011, 156869.

Opacic, T., Stevens, C. & Tillmann, B. (2009) Unspoken knowledge: Implicit learning of structured human dance movement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 35 (6), 1570–1577.

Opitz, B. & Friederici, A.D. (2004) Brain Correlates of Language Learning: The Neuronal Dissociation of Rule-Based versus Similarity-Based Learning. *Journal of Neuroscience*. 24 (39), 8436–8440.

Opitz, B. & Friederici, A.D. (2003) Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *NeuroImage*. 19 (4), 1730–1737.

Orlov, T., Yakovlev, V., Hochstein, S. & Zohary, E. (2000) Macaque monkeys categorize images by their ordinal number. *Nature*. 404 (6773), 77–80.

Ororbia, A. (2019) Spiking Neural Predictive Coding for Continual Learning from Data Streams. *arXiv*. 1908.08655.

O'Rourke, P. (2013) 'The interaction of different working memory mechanisms and sentence processing: A study of the P600', in *Proceedings of the Annual Meeting of the Cognitive Science Society*. [Online]. pp. 1097–1102.

Orpella, J., Ripollés, P., Ruzzoli, M., Amengual, J.L., Callejas, A., Martinez-Alvarez, A., Soto-Faraco, S. & Diego-Balaguer, R. de (2020) Integrating when and what information in the left parietal lobe allows language rule generalization. *PLOS Biology*. 18 (11), e3000895.

Osterhout, L. & Holcomb, P.J. (1992) Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*. 31 (6), 785–806.

Park, S.A., Miller, D.S., Nili, H., Ranganath, C. & Boorman, E.D. (2020) Map Making: Constructing, Combining, and Inferring on Abstract Cognitive Maps. *Neuron*. 107 (6), 1226-1238.e8.

Parr, T. & Friston, K.J. (2018) The Discrete and Continuous Brain: From Decisions to Movement—And Back Again. *Neural Computation*. 30 (9), 2319–2347.

Parthasarathy, A., Herikstad, R., Bong, J.H., Medina, F.S., Libedinsky, C. & Yen, S.-C. (2017) Mixed selectivity morphs population codes in prefrontal cortex. *Nature Neuroscience*. 20 (12), 1770–1779.

Pearson, K. (1901) LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 2 (11), 559–572.

Peña, M., Bonatti, L.L., Nespor, M. & Mehler, J. (2002) Signal-Driven Computations in Speech Processing. *Science*. 298 (5593), 604–607.

Peñaloza, C., Benetello, A., Tuomiranta, L., Heikius, I.-M., Järvinen, S., Majos, M.C., Cardona, P., Juncadella, M., Laine, M., Martin, N. & Rodríguez-Fornells, A. (2015) Speech segmentation in aphasia. *Aphasiology*. 29 (6), 724–743.

Peter, M.S. & Rowland, C.F. (2019) Aligning Developmental and Processing Accounts of Implicit and Statistical Learning. *Topics in Cognitive Science*. 11 (3), 555–572.

Peterfreund, E. & Gavish, M. (2021) Multidimensional scaling of noisy high dimensional data. *Applied and Computational Harmonic Analysis*. 51, 333–373.

Petersson, K.-M., Folia, V. & Hagoort, P. (2012) What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*. 120 (2), 83–95.

Petersson, K.M., Forkstam, C. & Ingvar, M. (2004) Artificial syntactic violations activate Broca's region. *Cognitive Science*. 28 (3), 383–407.

Petkov, C.I. & ten Cate, C. (2020) Structured Sequence Learning: Animal Abilities, Cognitive Operations, and Language Evolution. *Topics in Cognitive Science*. 12 (3), 828–842.

Petkov, C.I. & Jarvis, E.D. (2012) Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*. 4, 12.

Petkov, C.I. & Wilson, B. (2012) On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 367 (1598), 2077–2088.

Petrides, M. (1991) Functional specialization within the dorsolateral frontal cortex for serial order memory. *Proceedings of the Royal Society of London. Series B: Biological Sciences*. 246 (1317), 299–306.

Petrides, M. & Pandya, D.N. (2009) Distinct Parietal and Temporal Pathways to the Homologues of Broca's Area in the Monkey. *PLOS Biology*. 7 (8), e1000170.

Piai, V., Anderson, K.L., Lin, J.J., Dewar, C., Parvizi, J., Dronkers, N.F. & Knight, R.T. (2016) Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proceedings of the National Academy of Sciences*. 113 (40), 11366–11371.

Pinker, S. & Jackendoff, R. (2005) The faculty of language: what's special about it? *Cognition*. 95 (2), 201–236.

Plate, T.A. (1992) 'Holographic recurrent networks', in *Proceedings of the 5th International Conference on Neural Information Processing Systems*. NIPS'92. [Online]. 30 November 1992 San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. pp. 34–41.

Plate, T.A. (1995) Holographic reduced representations. *IEEE transactions on neural networks*. 6 (3), 623–641.

Poggio, T. (2012) The Levels of Understanding framework, revised. *Perception*. 41 (9), 1017–1023.

Poletiek, F.H., Conway, C.M., Ellefson, M.R., Lai, J., Bocanegra, B.R. & Christiansen, M.H. (2018) Under What Conditions Can Recursion Be Learned? Effects of Starting Small in Artificial Grammar Learning of Center-Embedded Structure. *Cognitive Science*. 42 (8), 2855–2889.

Poletiek, F.H., Monaghan, P., van de Velde, M. & Bocanegra, B.R. (2021) The semantics-syntax interface: Learning grammatical categories and hierarchical syntactic structure through semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 47 (7), 1141–1155.

Pothos, E.M. & Kirk, J. (2004) Investigating learning deficits associated with dyslexia. *Dyslexia (Chichester, England)*. 10 (1), 61–76.

Price, A.R., Peelle, J.E., Bonner, M.F., Grossman, M. & Hamilton, R.H. (2016) Causal Evidence for a Mechanism of Semantic Integration in the Angular Gyrus as Revealed by High-Definition Transcranial Direct Current Stimulation. *Journal of Neuroscience*. 36 (13), 3829–3838.

Price, C.J. (2010) The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*. 1191, 62–88.

Procyk, E., Ford Dominey, P., Amiez, C. & Joseph, J.P. (2000) The effects of sequence structure and reward schedule on serial reaction time learning in the monkey. *Brain Research. Cognitive Brain Research*. 9 (3), 239–248.

Prystauka, Y. & Lewis, A.G. (2019) The power of neural oscillations to inform sentence comprehension: A linguistic perspective. *Language and Linguistics Compass*. 13 (9), e12347.

Pullum, G.K. & Scholz, B.C. (2010) 'Recursion and the infinitude claim', in Harry van der Hulst (ed.) *Recursion and Human Language*. Studies in Generative Grammar. Berlin, Germany: De Gruyter Mouton. pp. 113–137.

Pulvermüller, F. & Fadiga, L. (2010) Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*. 11 (5), 351–360.

Rabiner, L. & Juang, B. (1986) An introduction to hidden Markov models. *IEEE ASSP Magazine*. 3 (1), 4–16.

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

Reber, A.S. (1967) Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*. 6 (6), 855–863.

Regel, S., Meyer, L. & Gunter, T.C. (2014) Distinguishing Neurocognitive Processes Reflected by P600 Effects: Evidence from ERPs and Neural Oscillations. *PLOS ONE*. 9 (5), e96840.

Restle, F. & Brown, E.R. (1970) Serial pattern learning. *Journal of Experimental Psychology*. 83 (1, Pt.1), 120–125.

Reynolds, J.H. & Heeger, D.J. (2009) The normalization model of attention. *Neuron*. 61 (2), 168–185.

Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K. & Fusi, S. (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature*. 497 (7451), 585–590.

Rizzolatti, G. & Arbib, M.A. (1998) Language within our grasp. *Trends in Neurosciences*. 21 (5), 188–194.

Roach, B.J. & Mathalon, D.H. (2008) Event-Related EEG Time-Frequency Analysis: An Overview of Measures and An Analysis of Early Gamma Band Phase Locking in Schizophrenia. *Schizophrenia Bulletin*. 34 (5), 907–926.

Roberts, B.M., Libby, L.A., Inhoff, M.C. & Ranganath, C. (2018) Brain activity related to working memory for temporal order and object information. *Behavioural Brain Research*. 354, 55–63.

Robertson, E.M. (2007) The Serial Reaction Time Task: Implicit Motor Skill Learning? *Journal of Neuroscience*. 27 (38), 10073–10075.

Rogalsky, C. & Hickok, G. (2009) Selective Attention to Semantic and Syntactic Features Modulates Sentence Processing Networks in Anterior Temporal Cortex. *Cerebral Cortex (New York, NY)*. 19 (4), 786–796.

Rogalsky, C., Matchin, W. & Hickok, G. (2008) Broca's Area, Sentence Comprehension, and Working Memory: An fMRI Study. *Frontiers in Human Neuroscience*. 2, 14.

Rogers, E., Mehta, S., Shengelia, R. & Reid, M.C. (2013) Four Strategies for Managing Opioid-Induced Side Effects in Older Adults. *Clinical geriatrics*. 21 (4).

Rohrmeier, M., Fu, Q. & Dienes, Z. (2012) Implicit Learning of Recursive Context-Free Grammars. *PLOS ONE*. 7 (10), e45885.

Rolls, E.T. & Treves, A. (2011) The neuronal encoding of information in the brain. *Progress in Neurobiology*. 95 (3), 448–490.

Romberg, A.R. & Saffran, J.R. (2013) All Together Now: Concurrent Learning of Multiple Structures in an Artificial Language. *Cognitive Science*. 37 (7), 1290–1320.

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

Rong, F., Isenberg, A.L., Sun, E. & Hickok, G. (2018) The neuroanatomy of speech sequencing at the syllable level. *PLoS ONE*. 13 (10), e0196381.

Rosenbaum, D.A., Cohen, R.G., Jax, S.A., Weiss, D.J. & van der Wel, R. (2007) The problem of serial order in behavior: Lashley's legacy. *Human Movement Science*. 26 (4), 525–554.

Rouault, M. & Koechlin, E. (2018) Prefrontal function and cognitive control: from action to language. *Current Opinion in Behavioral Sciences*. 21, 106–111.

Roy, A. (2017) The Theory of Localist Representation and of a Purely Abstract Cognitive System: The Evidence from Cortical Columns, Category Cells, and Multisensory Neurons. *Frontiers in Psychology*. 8, 186.

Rubin, J., Ulanovsky, N., Nelken, I. & Tishby, N. (2016) The Representation of Prediction Error in Auditory Cortex. *PLOS Computational Biology*. 12 (8), e1005058.

Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F. & Cushman, F. (2008) Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*. 107 (2), 479–500.

Saffran, J.R. (2002) Constraints on Statistical Language Learning. *Journal of Memory and Language*. 47 (1), 172–196.

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

Saffran, J.R., Johnson, E.K., Aslin, R.N. & Newport, E.L. (1999) Statistical learning of tone sequences by human infants and adults. *Cognition*. 70 (1), 27–52.

Saffran, J.R., Newport, E.L. & Aslin, R.N. (1996) Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*. 35 (4), 606–621.

Saffran, J.R., Newport, E.L., Aslin, R.N., Tunick, R.A. & Barrueco, S. (1997) Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*. 8 (2), 101–105.

Sahin, N.T., Pinker, S., Cash, S.S., Schomer, D. & Halgren, E. (2009) Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science (New York, N.Y.)*. 326 (5951), 445–449.

Sainburg, T., Theilman, B., Thielk, M. & Gentner, T.Q. (2019) Parallels in the sequential organization of birdsong and human speech. *Nature Communications*. 10 (1), 1–11.

Salinsky, M.C., Oken, B.S. & Binder, L.M. (1996) Assessment of drowsiness in epilepsy patients receiving chronic antiepileptic drug therapy. *Epilepsia*. 37 (2), 181–187.

Samsonovich, A.V. (2010) 'Toward a Unified Catalog of Implemented Cognitive Architectures', in *Biologically Inspired Cognitive Architectures 2010: Proceedings of the First Annual Meeting of the BICA Society*. [Online]. IOS Press. pp. 195–244.

Sato, N. & Yamaguchi, Y. (2010) Simulation of Human Episodic Memory by Using a Computational Model of the Hippocampus. *Advances in Artificial Intelligence*. 2010, e392868.

Schafer, E.W.P. & Marcus, M.M. (1973) Self-Stimulation Alters Human Sensory Brain Responses. *Science*. 181 (4095), 175–177.

Schapiro, A.C., Gregory, E., Landau, B., McCloskey, M. & Turk-Browne, N.B. (2014) The Necessity of the Medial Temporal Lobe for Statistical Learning. *Journal of Cognitive Neuroscience*. 26 (8), 1736–1747.

Schapiro, A.C., Turk-Browne, N.B., Botvinick, M.M. & Norman, K.A. (2017) Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Phil. Trans. R. Soc. B*. 372 (1711), 20160049.

Schiff, R., Sasson, A., Star, G. & Kahta, S. (2017) The role of feedback in implicit and explicit artificial grammar learning: a comparison between dyslexic and non-dyslexic adults. *Annals of Dyslexia*. 67 (3), 333–355.

Schiffer, A.-M., Ahlheim, C., Wurm, M.F. & Schubotz, R.I. (2012) Surprised at All the Entropy: Hippocampal, Caudate and Midbrain Contributions to Learning from Prediction Errors. *PLOS ONE*. 7 (5), e36445.

Schneider, J.M. & Maguire, M.J. (2018) Identifying the relationship between oscillatory dynamics and event-related responses. *International Journal of Psychophysiology*. 133, 182–192.

Schoenemann, P.T. (2009) Evolution of Brain and Language. *Language Learning*. 59 (s1), 162–186.

Schuck, N.W. & Niv, Y. (2019) Sequential Replay of Non-spatial Task States in the Human Hippocampus. *Science (New York, N.Y.)*. 364 (6447), eaaw5181.

Schwarb, H. & Schumacher, E.H. (2012) Generalized lessons about sequence learning from the study of the serial reaction time task. *Advances in Cognitive Psychology*. 8 (2), 165–178.

Segaert, K., Mazaheri, A. & Hagoort, P. (2018) Binding language: structuring sentences through precisely timed oscillatory mechanisms. *European Journal of Neuroscience*. 48 (7), 2651–2662.

Seitz, A.R., Kim, R., van Wassenhove, V. & Shams, L. (2007) Simultaneous and Independent Acquisition of Multisensory and Unisensory Associations. *Perception*. 36 (10), 1445–1453.

Sense, F. & Rijn, H. van (2018) Probabilistic motor sequence learning in a virtual reality serial reaction time task. *PLOS ONE*. 13 (6), e0198759.

Seth, A.K., Barrett, A.B. & Barnett, L. (2015) Granger Causality Analysis in Neuroscience and Neuroimaging. *Journal of Neuroscience*. 35 (8), 3293–3297.

Shahbaba, B., Li, L., Agostinelli, F., Saraf, M., Elias, G.A., Baldi, P. & Fortin, N.J. (2019) Hippocampal ensembles represent sequential relationships among discrete nonspatial events. *bioRxiv*. 840199.

Sharma, S., Voelker, A. & Eliasmith, C. (2017) 'A Spiking Neural Bayesian Model of Life Span Inference', in *Proceedings of the Annual Meeting of the Cognitive Science Society*. pp. 3131–3136.

Shastri, L. & Ajjanagadde, V. (1993) From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*. 16 (3), 417–451.

Shattuck, D.W. & Leahy, R.M. (2002) BrainSuite: an automated cortical surface identification tool. *Medical Image Analysis*. 6 (2), 129–142.

Sherfey, J.S., Soplata, A.E., Ardid, S., Roberts, E.A., Stanley, D.A., Pittman-Polletta, B.R. & Kopell, N.J. (2018) DynaSim: A MATLAB Toolbox for Neural Modeling and Simulation. *Frontiers in Neuroinformatics*. 12, 10.

Shimbo, A., Izawa, E.-I. & Fujisawa, S. (2021) Scalable representation of time in the hippocampus. *Science Advances*. 7 (6), eabd7013.

Shohamy, D. & Wagner, A.D. (2008) Integrating Memories in the Human Brain: Hippocampal–Midbrain Encoding of Overlapping Events. *Neuron*. 60 (2), 378–389.

Silva, S., Folia, V., Hagoort, P. & Petersson, K.M. (2017) The P600 in Implicit Artificial Grammar Learning. *Cognitive Science*. 41 (1), 137–157.

Silver, R.A. (2010) Neuronal arithmetic. *Nature reviews. Neuroscience*. 11 (7), 474–489.

Singer, W. (1999) Neuronal Synchrony: A Versatile Code for the Definition of Relations? *Neuron*. 24 (1), 49–65.

Smith, E.E. & Jonides, J. (1999) Storage and Executive Processes in the Frontal Lobes. *Science*. 283 (5408), 1657–1661.

Smolensky, P. (1990) Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*. 46 (1), 159–216.

Smolensky, P. & Legendre, G. (2011) *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar (Volume I: Cognitive Architecture)*. Reprint edition. Cambridge, Mass.; London: MIT Press.

Solomon, E.A., Lega, B.C., Sperling, M.R. & Kahana, M.J. (2019) Hippocampal theta codes for distances in semantic and temporal spaces. *Proceedings of the National Academy of Sciences*. 116 (48), 24343–24352.

Song, S., Howard, J.H. & Howard, D.V. (2008) Perceptual sequence learning in a serial reaction time task. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*. 189 (2), 145–158.

Sonnweber, R., Ravignani, A. & Fitch, W.T. (2015) Non-adjacent visual dependency learning in chimpanzees. *Animal Cognition*. 18 (3), 733–745.

Spear, D.J., Hienz, R.D. & Brady, J.V. (1992) Acute opioid administration effects on sensory and motor function in baboons: buprenorphine, morphine, and naloxone. *Behavioural Pharmacology*. 3 (1), 31–42.

Spierings, M.J. & ten Cate, C. (2016) Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. *Proceedings of the National Academy of Sciences*. 113 (27), E3977–E3984.

Stabler, E.P. (2004) Varieties of crossing dependencies: structure dependence and mild context sensitivity. *Cognitive Science*. 28 (5), 699–720.

Stachenfeld, K.L., Botvinick, M.M. & Gershman, S.J. (2017) The hippocampus as a predictive map. *Nature Neuroscience*. 20 (11), 1643–1653.

Stewart, T.C., Tripp, B. & Eliasmith, C. (2009) Python Scripting in the Nengo Simulator. *Frontiers in Neuroinformatics*. 3, 7.

Stobbe, N., Westphal-Fitch, G., Aust, U. & Fitch, W.T. (2012) Visual artificial grammar learning: comparative research on humans, kea (Nestor notabilis) and pigeons (Columba livia). *Philosophical Transactions of the Royal Society B: Biological Sciences*. 367 (1598), 1995–2006.

Summerfield, C. & de Lange, F.P. (2014) Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews. Neuroscience*. 15 (11), 745–756.

Sur, S. & Sinha, V.K. (2009) Event-related potential: An overview. *Industrial Psychiatry Journal*. 18 (1), 70–73.

Tabullo, Á., Sevilla, Y., Segura, E., Zanutto, S. & Wainselboim, A. (2013) An ERP study of structural anomalies in native and semantic free artificial grammar: Evidence for shared processing mechanisms. *Brain Research*. 1527, 149–160.

Takács, Á., Kóbor, A., Kardos, Z., Janacsek, K., Horváth, K., Beste, C. & Nemeth, D. (2021) Neurophysiological and functional neuroanatomical coding of statistical and deterministic rule information during sequence learning. *Human Brain Mapping*. 42 (10), 3182–3201.

Takahashi, S. (2013) Hierarchical organization of context in the hippocampal episodic code. Howard Eichenbaum (ed.). *eLife*. 2, e00321.

Tallon-Baudry, C., Bertrand, O., Delpuech, C. & Pernier, J. (1996) Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*. 16 (13), 4240–4249.

Tamalunas, A. (2021) *Chapter 3: Standardizing Data*. [Online]. Available from: https://bookdown.org/anta8363/fluoR_bookdown/stand.html (Accessed 22 November 2021).

Tanner, D. (2019) Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach. *Cortex*. 111, 210–237.

Tanner, D. & Van Hell, J.G. (2014) ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*. 56, 289–301.

Tavares, R.M., Mendelsohn, A., Grossman, Y., Williams, C.H., Shapiro, M., Trope, Y. & Schiller, D. (2015) A Map for Social Navigation in the Human Brain. *Neuron*. 87 (1), 231–243.

Teng, X., Tian, X. & Poeppel, D. (2016) Testing multi-scale processing in the auditory system. *Scientific Reports*. 6 (1), 34390.

Terry, J., Stevens, C.J., Weidemann, G. & Tillmann, B. (2016) Implicit learning of between-group intervals in auditory temporal structures. *Attention, Perception, & Psychophysics*. 78 (6), 1728–1743.

Tettamanti, M. & Perani, D. (2012) 'The neurobiology of structure-dependency in natural language grammar', in *The Handbook of the Neuropsychology of Language, 2 Volume Set*. [Online]. John Wiley & Sons. pp. 229–251.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Doyne Farmer, J. (1992) Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*. 58 (1), 77–94.

Theves, S., Fernandez, G. & Doeller, C.F. (2019) The Hippocampus Encodes Distances in Multidimensional Feature Space. *Current Biology*. 29 (7), 1226-1231.e3.

Theves, S., Neville, D.A., Fernández, G. & Doeller, C.F. (2021) Learning and Representation of Hierarchical Concepts in Hippocampus and Prefrontal Cortex. *Journal of Neuroscience*. 41 (36), 7675–7686.

Thompson, H.E., Robson, H., Lambon Ralph, M.A. & Jefferies, E. (2015) Varieties of semantic 'access' deficit in Wernicke's aphasia and semantic aphasia. *Brain*. 138 (12), 3776–3792.

Todaro, C., Marzetti, L., Valdés Sosa, P.A., Valdés-Hernandez, P.A. & Pizzella, V. (2019) Mapping Brain Activity with Electrocorticography: Resolution Properties and Robustness of Inverse Solutions. *Brain Topography*. 32 (4), 583–598.

Todorovic, A. & Lange, F.P. de (2012) Repetition Suppression and Expectation Suppression Are Dissociable in Time in Early Auditory Evoked Fields. *Journal of Neuroscience*. 32 (39), 13389–13395.

Torgerson, W.S. (1952) Multidimensional scaling: I. Theory and method. *Psychometrika*. 17 (4), 401–419.

Torgerson, W.S. (1958) *Theory and methods of scaling*. Oxford, England: Wiley.

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

Tosh, C.R. & Ruxton, G.D. (eds.) (2010) *Modelling Perception with Artificial Neural Networks*. Cambridge, England: Cambridge University Press.

Trujillo, O. & Eliasmith, C. (2014) A spiking-neuron model of memory encoding and replay in hippocampus. *BMC Neuroscience*. 15 (Suppl 1), P166.

Tsuchiya, N., Kawasaki, H., Oya, H., Iii, M.A.H. & Adolphs, R. (2008) Decoding Face Information in Time, Frequency and Space from Direct Intracranial Recordings of the Human Brain. *PLOS ONE*. 3 (12), e3892.

Uddén, J. & Bahlmann, J. (2012) A rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 367 (1598), 2023–2032.

Udden, J. & Männel, C. (2018) 'Artificial grammar learning and its neurobiology in relation to language processing and development', in *The Oxford Handbook of Psycholinguistics*. Oxford University Press. pp. 755–783.

van der Velde, F. & de Kamps, M. (2006) Neural blackboard architectures of combinatorial structures in cognition. *The Behavioral and Brain Sciences*. 29 (1), 37–70; discussion 70-108.

de Vries, M., Christiansen, M. & Petersson, K.M. (2011) Learning Recursion: Multiple Nested and Crossed Dependencies. *Biolinguistics*. 5 (1–2), 10–35.

Vuong, L.C., Meyer, A.S. & Christiansen, M.H. (2016) Concurrent Statistical Learning of Adjacent and Nonadjacent Dependencies. *Language Learning*. 66 (1), 8–30.

Wada, J. (1949) A new method of determining the side of cerebral speech dominance : A preliminary report on the intracarotid injection of sodium amytal in man. *Igaku to Seibutsugaku*. 14, 221–222.

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N. & Diedrichsen, J. (2016) Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*. 137, 188–200.

Wang, J., Yang, Y., Zhao, X., Zuo, Z. & Tan, L.-H. (2020) Evolutional and developmental anatomical architecture of the left inferior frontal gyrus. *NeuroImage*. 222, 117268.

Wang, L., Jensen, O., Brink, D. van den, Weder, N., Schoffelen, J.-M., Magyari, L., Hagoort, P. & Bastiaansen, M. (2012) Beta oscillations relate to the N400m during language comprehension. *Human Brain Mapping*. 33 (12), 2898–2912.

Wang, L., Uhrig, L., Jarraya, B. & Dehaene, S. (2015) Representation of Numerical and Sequential Patterns in Macaque and Human Brains. *Current Biology*. 25 (15), 1966–1974.

Wang, Y.-C., Grewal, S.S., Goyal, A., Alvi, M.A., Worrell, G., Brinkmann, B., Wong-Kisiel, L., Britton, J., Marsh, W.R., Burkholder, D., Payne, E., Shin, C., Cascino, G., Lundstrom, B.N., Wu, M.-H. & Gompel, J.J.V. (2020) Comparison of narcotic pain control between stereotactic electrocorticography and subdural grid implantation. *Epilepsy & Behavior*. 103 (Pt A), 106843.

Watson, S.K., Burkart, J.M., Schapiro, S.J., Lambeth, S.P., Mueller, J.L. & Townsend, S.W. (2020) Nonadjacent dependency processing in monkeys, apes, and humans. *Science Advances*. 6 (43), eabb0725.

Weber-Fox, C.M. & Neville, H.J. (1996) Maturational Constraints on Functional Specializations for Language Processing: ERP and Behavioral Evidence in Bilingual Speakers. *Journal of Cognitive Neuroscience*. 8 (3), 231–256.

Wernicke, C. (1874) *Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis*. Cohn.

Wernicke, C. (1881) *Lehrbuch der Gehirnkrankheiten für Aerzte und Studirende*. Fischer.

Werning, M., Hinzen, W. & Machery, E. (2012) *The Oxford Handbook of Compositionality*. OUP Oxford.

Whitney, H. (1936) Differentiable Manifolds. *Annals of Mathematics*. 37 (3), 645–680.

Whittington, J.C.R., Muller, T.H., Mark, S., Chen, G., Barry, C., Burgess, N. & Behrens, T.E.J. (2020) The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell*. 183 (5), 1249-1263.e23.

Whittington, M.A., Cunningham, M.O., LeBeau, F.E.N., Racca, C. & Traub, R.D. (2011) Multiple origins of the cortical gamma rhythm. *Developmental Neurobiology*. 71 (1), 92–106.

Whittle, P. (1963) On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*. 50 (1–2), 129–134.

Wiener, N. (1956) 'The theory of prediction', in E. F. Beckenbach (ed.) *Modern Mathematics for the Engineer*. New York, NY, USA: McGraw-Hill.

Williams, J.T., Darcy, I. & Newman, S.D. (2015) Modality-independent neural mechanisms for novel phonetic processing. *Brain Research*. 1620, 107–115.

Willmore, B.D.B. & King, A.J. (2009) Auditory Cortex: Representation through Sparsification? *Current biology : CB*. 19 (24), R1123–R1125.

Wilson, B., Kikuchi, Y., Sun, L., Hunter, D., Dick, F., Smith, K., Thiele, A., Griffiths, T.D., Marslen-Wilson, W.D. & Petkov, C.I. (2015) Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nature Communications*. 6, 8901.

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

Wilson, B., Slater, H., Kikuchi, Y., Milne, A.E., Marslen-Wilson, W.D., Smith, K. & Petkov, C.I. (2013) Auditory artificial grammar learning in macaque and marmoset monkeys. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*. 33 (48), 18825–18835.

Wilson, B., Smith, K. & Petkov, C.I. (2015) Mixed-complexity artificial grammar learning in humans and macaque monkeys: evaluating learning strategies. *European Journal of Neuroscience*. 41 (5), 568–578.

Wilson, B., Spierings, M., Ravignani, A., Mueller, J.L., Mintz, T.H., Wijnen, F., Kant, A. van der, Smith, K. & Rey, A. (2018) Non-adjacent Dependency Learning in Humans and Other Animals. *Topics in Cognitive Science*. 12 (3), 843–858.

Wilson, W.H., Halford, G.S., Gray, B. & Phillips, S. (2001) 'The STAR-2 Model for Mapping Hierarchically Structured Analogs', in *World Bank, Human Development 4 (AFTH4). Washington DC*. [Online]. 2001 MIT Press. pp. 125–159.

Witt, A. & Vinter, A. (2012) Artificial grammar learning in children: abstraction of rules or sensitivity to perceptual features? *Psychological Research*. 76 (1), 97–110.

Wixted, J.T., Squire, L.R., Jang, Y., Papesh, M.H., Goldinger, S.D., Kuhn, J.R., Smith, K.A., Treiman, D.M. & Steinmetz, P.N. (2014) Sparse and distributed coding of episodic memory in neurons of the human hippocampus. *Proceedings of the National Academy of Sciences*. 111 (26), 9621–9626.

Wu, C.-Y., Vissiennon, K., Friederici, A.D. & Brauer, J. (2016) Preschoolers' brains rely on semantic cues prior to the mastery of syntax during sentence comprehension. *NeuroImage*. 126, 256–266.

Yan, C., Christophel, T.B., Allefeld, C. & Haynes, J.-D. (2021) Decoding verbal working memory representations of Chinese characters from Broca's area. *NeuroImage*. 226, 117595.

Yi, H.G., Leonard, M.K. & Chang, E.F. (2019) The Encoding of Speech Sounds in the Superior Temporal Gyrus. *Neuron*. 102 (6), 1096–1110.

Yu, A.C. & Margoliash, D. (1996) Temporal hierarchical control of singing in birds. *Science (New York, N.Y.)*. 273 (5283), 1871–1875.

Zalesak, M. & Heckers, S. (2009) The role of the hippocampus in transitive inference. *Psychiatry research*. 172 (1), 24–30.

Zeithamova, D., Schlichting, M.L. & Preston, A.R. (2012) The hippocampus and inferential reasoning: building memories to navigate future decisions. *Frontiers in Human Neuroscience*. 6, 70.

Zhuang, P., Dang, N., Waziri, A., Gerloff, C., Cohen, L.G., Hallett, M. & Warzeri, A. (1998) Implicit and explicit learning in an auditory serial reaction time task. *Acta Neurologica Scandinavica*. 97 (2), 131–137.

Zielinski, M.C., Shin, J.D. & Jadhav, S.P. (2019) Coherent Coding of Spatial Position Mediated by Theta Oscillations in the Hippocampus and Prefrontal Cortex. *Journal of Neuroscience*. 39 (23), 4550–4565.

Zilles, K. & Amunts, K. (2018) Cytoarchitectonic and receptorarchitectonic organization in Broca's region and surrounding cortex. *Current Opinion in Behavioral Sciences*. 21, 93–105.