Musical Context Facilitates Event Segmentation and Sequential Learning Through Interconnected Neural Networks and Strengthened Hippocampal Encoding

A Dissertation
Presented to
The Academic Faculty

by

Yiren Ren

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Psychology, College of Sciences

Georgia Institute of Technology
[December 2024]

COPYRIGHT © 2024 BY YIREN REN

Musical Context Facilitates Event Segmentation and Sequential Learning Through Interconnected Neural Networks and Strengthened Hippocampal Encoding

Approved by:

Dr. Thackery Brown, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Lila Davachi
Department of Psychology
Columbia University

Dr. Paul Verhaeghen
School of Psychology
Georgia Institute of Technology

Dr. Elizabeth Race
Department of Psychology
Tufts University

Dr. Sashank Varma
School of Psychology
Georgia Institute of Technology

Date Approved: [Nov 18, 2024]

To families, friends, colleagues and my participants

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to everyone who supported me throughout this journey. My supervisor, Thackery, has been an incredible mentor. Thank you for encouraging my innovative spirit and trusting me to explore new paths in my research. I'm deeply grateful to Dr. Vishwadeep Ahluwalia for his patient guidance and expertise in MRI analysis. To my family, your love and care have been a constant source of strength, despite the vast distance between us. I also owe special thanks to my furry buddy, Senbei, who has been the best emotional support cat a PhD student can have. Finally, to my husband, Sean: your unwavering support has been the bedrock of this achievement. Thank you for being my biggest fan and for believing in me every step of the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

MTL	Medial Temporal Lobe
vmPFC	Ventral Medial Prefrontal Cortex
NAc	Nucleus Accumbens
PHC	Parahippocampal Cortex
aPHC	Anterior Parahippocampal Cortex
pPHC	Posterior Parahippocampal Cortex
IFG	Inferior Frontal Gyrus
SFG	Superior Frontal Gyrus
MFG	Middle Frontal Gyrus
STG	Superior Temporal Gyrus
AG	Angular Gyrus
LOC	Lateral Occipital Cortex
DG	Dentate Gyrus
RSM	Representational Similarity Matrix
RSA	Representational Similarity Analysis
FIR	Finite Impulse Response
ROI	Regions of interest

SUMMARY

Temporal order memory, the ability to recall the sequence of items or events, is a critical aspect of human cognition. It underpins key processes such as language comprehension, decision-making, and learning. Extracting patterns from continuous experiences helps individuals form an understanding of the temporal relationships between events, allowing for future predictions and adaptive behavior. Prior research highlights the role of context in shaping the encoding of the temporal order and distance between events. Music is a special event that can serve as a contextual backdrop for many memories, and importantly it has its own built-in sequence structure which can provide strong temporal cues to the listener. Thus it is plausible that music, when treated as a collection of contextual cues, enhances temporal order learning and modulates the neural mechanisms underlying this process. To test this, participants engaged in a statistical learning and event segmentation task, viewing streams of images to extract temporal patterns while undergoing fMRI scanning. Some of these streams were paired with familiar music, hypothesized to serve as a structured series of cues that "scaffolds" learning the temporal relationships of the visual information, while other streams were learned in silence. Replicating and extending my prior studies, behavioral results demonstrated that music facilitated both sequence learning and event boundary detection. Neuroimaging results revealed distinct activation patterns between music and control conditions, particularly within regions of interest in the medial temporal lobe (MTL), prefrontal cortex (PFC), and striatum. Functional connectivity analyses showed that music enhanced connectivity between the MTL and visual processing regions, as well as between the MTL and ventromedial prefrontal cortex (vmPFC). Additionally, representational similarity analysis (RSA) of hippocampal activity revealed that music modulated neural representations of item relationships both within and across event

boundaries. The findings suggest that music strengthens associative binding within sequences and enhances the distinction between event boundaries, improving learning efficiency. This research contributes novel insights into cross-modal influences on memory, offering potential applications in educational and clinical contexts.

CHAPTER 1. INTRODUCTION

1.1. Temporal Order Memory

Temporal order memory, the ability to accurately sequence items and events, is fundamental to numerous aspects of human cognition and behavior. This capacity underpins essential skills such as language comprehension, navigation, motor skill development, and autobiographical memory recall. Event segmentation involves identifying boundaries that define when one episodic memory starts and ends, such as changes in location, activity, or goal. These event boundaries help us chunk our experiences into discrete episodes, allowing us to pinpoint specific moments in the past during recall (Tulving, 2002). Temporal regularity extraction, on the other hand, involves identifying patterns and consistencies across different encoding events. By recognizing the typical structure and sequence of stimuli and events, we can group similar episodes together in memory, despite day-to-day variability in details. This process assists in forming schemas that enable us to predict future events and plan accordingly (Kurby & Zacks, 2008; Sherman et al., 2020). These learning processes involve not only identifying and recalling individual items or events but also comprehending and establishing positional and temporal associations among them. Thus, understanding the mechanisms of temporal order memory and exploring methods to enhance this cognitive function hold both theoretical significance and practical applicability across various domains of human life.

One widely used task format to investigate the acquisition of temporal event order is statistical learning paradigms, which typically involve constructing 'artificial grammars' in which “noisy” stimulus presentations actually have stable transitional probabilities between items. This type of study aims to explore how humans acquire temporal order through exposure to

regularities in environmental inputs (Saffran et al., 1996). Participants are usually exposed to continuous presentations of items and engage in tasks such as repetition detection (Turk-Browne et al., 2005), boundary identification (E. K. Johnson & Tyler, 2010), or passive perception (Schapiro et al., 2012). Importantly, while often conceptualized as an implicit learning process, explicit knowledge acquisition can also reveal statistical learning effects (Batterink et al., 2015; Franco et al., 2011). This approach has been applied across various modalities, including language (Romberg & Saffran, 2010; Saffran et al., 1996), music (Jarret et al., 2019; Tillmann & McAdams, 2004), motor sequences (Nissen & Bullemer, 1987; Schendan et al., 2003), visual shapes (Fiser & Aslin, 2002; Kirkham et al., 2002), and real-life scenarios (Brady & Oliva, 2008; Conway & Christiansen, 2006). To illustrate statistical learning using real-life example, when driving to work, you may often—but not always—pass the same neighbors and encounter similar traffic bottlenecks. Despite these events not occurring in the exact same place or time each day, it is useful (and we are able) to learn these regularities in sequences of events. Collectively, statistical learning studies demonstrate humans' innate capacity to automatically apprehend statistical relationships within temporal events, even for hierarchical regularities and at levels of generalization and abstraction (Aslin & Newport, 2012; Schapiro et al., 2014).

In service of both individual episodic memories but also “grouping” memories for the purposes of statistical learning, current literature has revealed that humans instinctively divide ongoing events into manageable chunks which facilitate enhanced processing and memory encoding (Kurby & Zacks, 2008; Zacks et al., 2007). This *Event Segmentation Theory* proposes that individuals refresh their memory representation, or 'event model,' at the juncture of transitioning events (Kurby & Zacks, 2008; Zacks et al., 2007). Through repetitive learning, this model becomes increasingly adept at forecasting forthcoming sequences precisely at these

boundaries (Radvansky & Zacks, 2017). To investigate event segmentation, researchers have employed various experimental tasks. These include asking participants to explicitly identify event boundaries in ongoing sequences (Zacks et al., 2001), examining how changes in context affect memory for temporal order (DuBrow & Davachi, 2013, 2014), and assessing how event boundaries impact the binding of information across time (Ezzyat & Davachi, 2014).

Central to statistical learning and event segmentation is the role of context, which is critical in defining where and when an episode occurred (Tulving, 1983; Watkins & Tulving, 1975). For temporal order memory, changes in context create event boundaries (DuBrow & Davachi, 2013). Such work from Davachi and colleagues demonstrated that changes in context can lead to separation in memory between events, affecting both the accuracy and the subjective experience of temporal order memory (Clewett et al., 2019; DuBrow & Davachi, 2013, 2016). Various contextual modifications have been used to create event boundaries, including changes in task types (DuBrow & Davachi, 2013), visual stimuli categories (DuBrow & Davachi, 2016), spatial locations (Pettijohn et al., 2016), background color (Wen & Egner, 2022), and reward prediction error (Rouhani et al., 2020). Many studies suggest better temporal order memory for within-event items than for items separated by contextual changes (DuBrow & Davachi, 2013; Heusser et al., 2018), highlighting that shared context helps bind items and estimate their temporal and ordinal distance. Consistent item-context associations, with context reinstatement during retrieval, can improve temporal order decisions and event segmentation abilities (Cox et al., 2021; Wen & Egner, 2022). This perspective suggests that context provides temporal cues for better determining the order of items. Moreover, consistency in context can improve sequential item learning and facilitate the generalization of temporal regularity into higher-level rules (Aslin & Newport, 2012).

Given our understanding of humans' abilities to cluster ongoing sequences of events and extract temporal patterns, an important question is: What factors influence the efficiency of temporal order memory, and is it possible to enhance the strength and clarity of temporal context traces on the formation of temporal order memory? Among various environmental factors that could potentially enhance temporal order memory, music emerges as a particularly promising candidate due to its inherent temporal organization. Music naturally contains multiple levels of temporal structure - from basic beat patterns to higher-level rhythmic phrases and overall musical form - making it a rich source of temporal cues. These hierarchical temporal patterns in music align closely with how humans naturally segment and organize temporal information in memory. Furthermore, music's temporal structure is highly predictable and consistent compared to many other environmental stimuli, potentially making it an effective scaffold for temporal memory formation. This inquiry is especially relevant for populations with memory impairments, such as those with Alzheimer's disease, who often struggle with encoding the order of events, leading to deficits in autobiographical memory, planning, and decision-making (Bellassen et al., 2012; D. L. Johnson & Kesner, 1997; Sullivan & Sagar, 1989). In this study, I focus on investigating how music's structured temporal patterns might serve as an effective tool for enhancing temporal order memory.

1.2. Music as a Contextual Cue for Memory and Current Gap

Music, omnipresent in human daily life, is a combination of sequences of multiple features including tones, rhythms, and timbres. Despite its intricate sequential nature, humans exhibit an incredible capacity to encode, recall, and even create music. While substantial literature has demonstrated music's role as a potent contextual cue aiding memory retrieval (El Haj et al., 2012; Ferreri et al., 2015; Smith, 1985; Wallace, 1994), less is known about how

background music during encoding influences learning performance with other stimuli, especially for non-verbal information which can't take on a "lyric-like" mnemonic relationship with the music.

In a previous study, I investigated whether familiar music characterized by high predictability could provide a temporal scaffold for the concurrent learning of visual sequences. This hypothesis was motivated by schema theories in memory literatures, which posit that new information linked to existing knowledge can be encoded more efficiently (e.g. van Kesteren et al., 2012). The data from my previous study suggested that music which was more predictable (either because the composition was veridically familiar, or the composition's structure was at least syntactically regular [following known music grammar]) could improve learning the order of pictures happening in parallel (Ren, Leslie, et al., 2024). Furthermore, I have shown neural evidence for strengthened cortical-hippocampal interactions during music-accompanied learning (under review), suggesting how potential information transfer between existing music memory structures and the novel visual sequential memory could occur. Indeed, previous studies have indicated that multisensory representations of statistical and probabilistic information can be transferrable across modalities (Glicksohn & Cohen, 2013; Mitchel & Weiss, 2011; Yildirim & Jacobs, 2015; D. Zhang et al., 2004). These findings suggest the potential for leveraging a schematic or familiar sequence, such as music, to facilitate the learning of statistical or temporal relationships in a different modality.

While extensive research exists on temporal order memory and statistical learning, significant gaps remain in our understanding of how these processes can be enhanced through cross-modal interactions. The present study addresses these gaps by investigating music's potential as a collection of contextual cues for improving visual sequence learning. First, most

statistical learning studies have focused on single modalities, such as auditory or visual inputs. However, real-life scenarios typically involve multi-modal stimuli, and music is a frequent, salient (Ren et al., 2024) co-occurrence with other stimuli which could scaffold or enhance learning in another modality. This study addresses this gap by investigating how music provides context signals which influence visual sequence learning and statistical learning. Second, although previous research has identified brain regions involved in temporal order memory and music processing separately, our understanding of whether music modulates the interactions between these neural systems in support of learning remains limited. This study aims to elucidate these neural dynamics, particularly focusing on brain network interactions during music-contextualized visual temporal order memory processing.

1.3. Background on Medial Temporal Lobe

Neuroscience research has identified a distributed brain network crucial for temporal order learning and memory. This network primarily involves three key regions: the medial temporal lobe (MTL), the prefrontal cortex (PFC) and the striatum. The medial temporal lobe (MTL), particularly the hippocampus, plays a central role in temporal order memory. It is crucial for detecting regular patterns in sequential stimuli and binding items to their temporal contexts (Eichenbaum, 2013). This region has been consistently implicated in items association and the extraction and encoding of temporal regularities, regardless of the sensory modality of the input. For instance, the hippocampus shows increased activity during repeated motifs in music listening tasks (Burunat et al., 2014) and gradually develops similar neural representations for adjacent items in visual statistical learning tasks (Schapiro et al., 2012). Animal studies have revealed related neuronal mechanisms, suggesting that hippocampal cells fire to code specific time and space, providing representations for a hierarchical memory map for a series of events(Howard &

Eichenbaum, 2015). The hippocampus is also essential for disambiguating overlapping sequences (Ranganath & Hsieh, 2016). For example, hippocampal neural patterns predicted which spatial sequence participants would choose before entering overlapping routes (Brown et al., 2016). This ability to distinguish shared sequences is critical for navigating complex environments (contexts) and recalling similar but distinct events.

Moreover, hippocampal neurons are sensitive to specific learned temporal positions of objects (Hsieh et al., 2014) and can distinguish ongoing sequences at community boundaries in hierarchical temporal sequences (Schapiro et al., 2016). The hippocampus also provides contextual representation and is sensitive to context changes (Maurer & Nadel, 2021), with it showing pattern separation when the environment changes, implementing the disambiguation between events even when they share overlaps (Leutgeb et al., 2007). On the other hand, the hippocampus's pattern similarity has been associated with the ability to recognize temporal order relationships between items especially when context changed (Ezzyat & Davachi, 2014), suggesting its role in binding discontinuous events (Wallenstein et al., 1998). Overall, the contextual sensitivity allows the hippocampus to segment continuous experiences as well as to bind distant events, a crucial function for organizing memories in temporal dimension and predicting future occurrences.

MTL involvement in temporal order memory is not limited to the hippocampus alone. The parahippocampal cortex, for instance, has been implicated in supporting the recollection of contextual information for temporal order memory in autobiographical events (St. Jacques et al., 2008). Furthermore, the MTL interacts with other brain regions, such as the prefrontal cortex and the striatum, to support complex temporal order learning and memory processes. For example, strong functional connectivity between the hippocampus and cerebellum has been observed

during music boundary segmentation tasks, positively correlating with section segmentation performance (Burunat et al., 2018). This aligns with evidence from visual serial sequence learning tasks showing that the hippocampal-cerebellum network contributes to spatio-temporal sequence encoding and prediction ability (Onuki et al., 2015). Taken altogether, the MTL's role in sequential memory encoding primarily relates to coding the temporal and spatial context and binding or dissociating items regarding their spatiotemporal relationship. Given this role, this study specifically examines how musical context might modulate MTL activity during visual sequence learning.

1.4. Background on Prefrontal Cortex

The prefrontal cortex (PFC) is the second region of interest that plays an important role in temporal order encoding. Much like the hippocampus, PFC cells exhibit firing patterns that reflect timing during learning and planning processes. The PFC has been implicated in maintaining the temporal information of visual stimuli during working memory tasks in primates and humans (Amiez & Petrides, 2007; Ninokura et al., 2003; Roberts et al., 2018). Furthermore, the PFC, in conjunction with the medial temporal lobe (MTL), is frequently implicated in supporting the encoding of associated events. Human fMRI data collected during item-specific learning and item relational information learning have suggested that the dorsolateral PFC plays a specific role in encoding relational information, while the ventrolateral PFC has a more general role in long-term memory formation (Blumenfeld et al., 2011; Murray & Ranganath, 2007). Moreover, beyond facilitating initial relational encoding, the PFC may interact with the MTL by projecting top-down temporal information from stored schemas or prior knowledge (mainly from the ventromedial PFC) to the MTL to enhance the encoding of new events which are associate in some form to the existing schematic memory structure (sometimes referred to as “resonance”

with past memory traces) (McAndrews & Milner, 1991; van Buuren et al., 2014; van Kesteren et al., 2012).

The PFC's engagement in temporal association learning is also sensitive to contextual factors. For instance, the medial PFC appears to be more sensitive to encoding events occurring within the same context (Davachi & DuBrow, 2015; Ezzyat & Davachi, 2011). In a hierarchical statistical learning task, medial PFC showed stronger connectivity with the hippocampus at event boundary when sequences changed context (Schapiro et al., 2016). During the serial recall of items and their temporal relationships, the lateral PFC demonstrated increased BOLD signal for cross-boundary items, while the medial PFC showed increased functional connectivity with the hippocampus when items were within the same context (Clewett et al., 2019). Specifically, the medial PFC plays a role in the formation and storage of contextual information, and the context-related representation in the medial PFC may project to the MTL to facilitate memory binding or retrieval under similar contexts (Polyn & Kahana, 2008; Preston & Eichenbaum, 2013). This prefrontal cortex-hippocampus circuit may be particularly critical in hierarchical sequential learning scenarios with changing contexts. In an earlier investigation of mine, data suggested stronger PFC-MTL connectivity during sequential learning in the presence of music, potentially indicating that the introduction of distinct and schematic music contexts enhances sequential encoding (Ren & Brown, under review). However, this prior study employed a controlled experimental paradigm where participants explicitly attended to both music and visual sequences. In real-world scenarios, people often encounter sequential information while background music is playing - for instance, learning a new route while listening to music, or observing patterns in the environment while music plays in a store. These naturalistic situations differ from laboratory conditions in two critical ways: First, the music-sequence associations are

less explicit, and second, the to-be-learned sequences contain more variability and noise, requiring statistical learning to extract underlying patterns. Statistical learning better reflects how humans acquire temporal knowledge in daily life, where perfect sequence repetition is rare and patterns must be discovered amid variable inputs. Therefore, while the earlier findings provide a foundation for the current work, it remains unknown whether music's facilitatory effects on sequence learning extend to these more ecologically valid contexts where patterns must be extracted from noisy input through statistical learning mechanisms.

Other than the medial PFC, of particular interest in the context of sequential learning and music processing is the inferior frontal gyrus (IFG). The IFG has been consistently highlighted in studies of sequential and statistical learning tasks across various modalities. In language processing, the IFG (particularly Broca's area) is crucial for syntactic processing and has been implicated in artificial grammar learning tasks (Petersson et al., 2012). Similarly, in music, the IFG is involved in processing harmonic structure and detecting violations in musical syntax (Sammler et al., 2009). This parallel suggests a domain-general role in sequential structure processing. In statistical learning, the IFG has been shown to be active during the acquisition of novel auditory and visual sequences. Karuza et al. (2013) found increased activation in the IFG during a word segmentation task that relied on statistical learning (Karuza et al., 2013). In visual statistical learning tasks, the IFG has been implicated in the extraction of temporal regularities from continuous visual streams (Turk-Browne et al., 2009). As a result, the IFG's dual role in both musical syntax and statistical learning makes it a key region of interest for investigating how musical context might facilitate visual sequence learning in the present study.

1.5. Background on The Striatum

The striatum plays a central role in learning timing and temporal patterns. Both animal and human studies have indicated that striatal cells exhibit regular firing patterns to support internal timing functions, such as relative timing coding (Mello et al., 2015) and interval duration representation (Matell et al., 2003). Furthermore, the striatum is involved in encoding temporal regularities. For example, the putamen has been implicated in supporting the encoding of rhythm regularities during music listening (Grahn & Rowe, 2009). While hippocampus is primarily involved in the extraction of regularities to associate items with their contexts (Ranganath, 2010), the striatum appears to provide predictions of temporal patterns during encoding (Grahn & Rowe, 2013). This function is thought to underlie the striatum's contribution to statistical learning tasks, such as probabilistic category learning (Aron et al., 2004) and visual segmentation (Turk-Browne et al., 2010). The striatum's role in temporal prediction and learning is closely tied to its dopaminergic mechanisms. The rewarding feedback resulting from the online comparison between predictions and inputs in the striatum might facilitate updates in the prefrontal cortex, enhancing the encoding of to-be-learned information and planning for subsequent cognitive behaviors (Dagher et al., 2001; van der Meer et al., 2010).

The striatum's connections with other brain regions are crucial for its role in temporal order memory. The prefrontal cortex, which receives prediction-related feedback from the striatum, may generate emotional motivation and stimulate the limbic system, including the hippocampus, to further encode the information (Haber & Knutson, 2010). This collaborative relationship has been implicated in tasks involving spatial navigation, which require both sequential encoding and movement planning (Goodroe et al., 2018). Furthermore, the striatum

and the hippocampus might collaborate to encode items within their temporal and spatial contexts, thereby enhancing declarative memory formation (Polti et al., 2022).

In the context of music, the striatum has been shown to play a role in rhythm perception and production. Studies have found that the putamen is particularly active during beat perception and synchronization tasks (Grahn & Brett, 2007). This suggests that the striatum's timing functions extend to the processing of musical rhythms, which could have implications for how music might influence temporal order learning in other domains. The striatum's involvement in reward processing also makes it relevant to music-related memory tasks. The experience of listening to pleasurable music has been associated with dopamine release in the striatum (Salimpoor et al., 2011), which could potentially enhance memory encoding processes. This reward-related activation could explain why music often serves as a powerful mnemonic device and why it might facilitate the encoding of temporal order information in other modalities.

1.6. Summary of Background

In summary, the MTL, PFC, and striatum form an interconnected network that supports temporal order memory through complementary functions. While each region contributes distinct computational processes - the MTL binding items to temporal contexts and detecting regularities, the PFC maintaining and manipulating temporal information while providing top-down schematic knowledge, and the striatum supporting temporal prediction and rhythm processing - their interactions are particularly relevant for music-facilitated learning. During music perception, the striatum's precise temporal processing of rhythmic patterns may provide a temporal scaffold that enhances the PFC's maintenance of sequential information. Simultaneously, the PFC may use existing musical schemas to guide the MTL's binding of novel visual sequences to the temporal structure provided by music. This may be especially effective

because music activates all three regions: engaging the striatum through rhythmic processing and reward, the MTL through contextual binding and pattern detection, and the PFC through structural analysis and schema activation. The coordination among these regions might explain why music could serve as an effective contextual framework for enhancing temporal order memory: it simultaneously provides temporal prediction (striatum), schematic structure (PFC), and contextual binding opportunities (MTL). Understanding how music modulates the interactions between these regions during visual sequence learning is therefore crucial for leveraging music's potential as a memory enhancement tool.

1.7. The Current Study and Hypotheses

The literature reviewed in the preceding sections highlights the critical roles of the medial temporal lobe (MTL), prefrontal cortex (PFC), and striatum in temporal order learning and memory. These regions work in concert to support the encoding, retrieval, and integration of sequential information across various modalities. However, questions remain about how these neural systems might be modulated by contextual factors, such as the presence of familiar, structured auditory sequences like music. This study seeks to build upon the previous research by investigating the impact of music on parallel sequential memory encoding using fMRI. Unlike the design from my earlier graduate program research (Ren et al., 2024) where participants deliberately learned individual sequences of four shapes at a time, the current study employs a statistical learning paradigm, which more-closely resembles real-life scenarios in which individuals continually encode sequences of events and the structure of the memory may not always be intentionally organized. Moreover, the study detailed in this dissertation presented music as a “background stimulus” as opposed to an intentional study tool the participants should use in learning. This extends my earlier research to again explore a more common memory

scenario from the real world (e.g., music on the radio while commuting to work each morning). These adaptations aim to better elucidate how music can facilitate temporal order and regularity learning, from the perspective of event segmentation processes.

Based on previous literature, I developed several specific hypotheses regarding both behavioral and neural outcomes. At the behavioral level, I predicted that music would enhance sequence learning, reflected by learning efficiency and subsequent retrieval. First, I hypothesized that sequences learned with musical context would show higher subsequent retrieval accuracy compared to sequences learned in silence, as music provides an additional contextual framework for encoding. Second, I predicted that event boundary detection during encoding would be more accurate when sequences were accompanied by music, as musical transitions could provide additional segmentation cues. Third, I expected that the sequence-music associations would benefit learning efficiency of the items sequential relationships, resulting in faster learning speed during encoding.

At the neural level, I formed three sets of hypotheses to explain the potential learning benefits of music. 1) 'Event Segmentation Effect': regarding event segmentation, I predicted that the medial temporal lobe (MTL) and inferior frontal gyrus (IFG) would show enhanced boundary-related activity in the music condition, reflecting more efficient event segmentation. This prediction is based on previous studies implicating these regions in responding to event boundaries and segmenting ongoing experiences (Burunat et al., 2024; Ezzyat & Davachi, 2014).

2) 'Network Organization Changes': concerning sequential relationship learning within the (boundaries of) sequences, I adopted a network-analysis approach in addition to examining activity patterns. While I expected the hippocampus and other MTL regions to be engaged in encoding temporal order based on prior literature (Hsieh et al., 2014; Ranganath & Hsieh, 2016),

I hypothesized that the network organization tied to learning would shift in the presence of music. This is because music provides a new contextual input from its associated cortical stream and might reduce computational demands on some of the circuitry relevant for visual sequence learning by strengthening and clarifying temporal contextual cues (Ren and Brown, under review). I predicted that the music condition would engage a broader brain network, particularly stronger functional connectivity within the prefrontal cortex (PFC), striatum, and MTL, reflecting the engagement of a music-schema-based learning mechanism (Preston & Eichenbaum, 2013; van Kesteren et al., 2012). Specifically, I expected to see higher correlation between the network patterns and the sequence learning performance. These connectivity patterns would indicate more coordinated neural processing and the integration of musical contextual information with visual sequences, facilitating more efficient encoding and consolidation of temporal order memory.

3) 'Hippocampal Representation Patterns': I was particularly interested in how music modulated hippocampal representations, given the hippocampus' critical role in associative binding and event separation (Ezzyat & Davachi, 2014; Wallenstein et al., 1998). I hypothesized that the presence of musical context would enhance these hippocampal functions, leading to higher pattern similarity for within-sequence items in the music condition. Conversely, I predicted that items across sequence boundaries would show greater pattern dissimilarity in the music condition, indicating enhanced event separation. These predictions would suggest that music modulates hippocampal representations to create more distinct neural patterns for different sequences while promoting coherence within sequences, ultimately supporting more efficient temporal order memory formation.

Additionally, I predicted that music would reduce the overall neural demand in brain areas associated with sequential and statistical learning, such as the angular gyrus (Van Opstal et al., 2008), lateral occipital cortex (Hsieh et al., 2011), and prefrontal cortex (Knutson et al., 2004). This hypothesis aligns with 1) the common assumption in fMRI literature that greater engagement tends to reflect greater processing demands on the brain region (Neubauer & Fink, 2009) and 2) extant evidence from my previous findings (Ren et al., 2024; neural data under review), which suggested that music can facilitate the learning process by decreasing activity in these regions.

1.8. Potential Implications and Significance

The intricate interplay among the medial temporal lobe, prefrontal cortex, and striatum in temporal order memory provides a neural framework for understanding music's influence on visual sequence learning. By investigating how these regions respond and interact differently when familiar music accompanies encoding, this study bridges the gap between auditory and visual statistical learning processes, addressing a significant question in cross-modal learning research.

This research uniquely combines a statistical learning paradigm with musical context, allowing us to examine how a familiar auditory sequence might scaffold the learning of novel visual sequences. By using fMRI to observe neural dynamics during this process, we can gain insight into how the brain integrates information across modalities to enhance temporal order memory.

The implications of this research extend beyond theoretical neuroscience into practical applications. Understanding music's modulatory effects on temporal order learning could inform the development of more effective educational strategies, particularly for teaching sequential

information. This knowledge could prove especially valuable for clinical populations with temporal processing deficits, such as individuals with neurodegenerative disorders or developmental conditions affecting executive function. Furthermore, this research may elucidate the neurobiological basis of music's effectiveness as a mnemonic device, potentially leading to optimized protocols for enhancing cognitive functions through musical scaffolding.

In the subsequent chapters, I present the experimental design, findings, and their contributions to understanding cross-modal influences on temporal order memory.

CHAPTER 2. METHOD

2.1. Participants

Eighty-nine healthy participants (42 males and 47 females) were recruited from the Georgia Institute of Technology's volunteer pool. Participants' ages range from 18 to 35, with an average of 20.57 years and standard deviation of 3.27 years old. Out of the 89 participants, 36 underwent fMRI scanning, while the remaining 53 participants completed the task with behavioral measures only. Among the 36 fMRI participants, 12 were males and 24 were females. These participants ranged from 18 years old to 35 years old (mean = 21.58, std = 4.1). However, one participant encountered technical issues during the MRI session, and stimuli timings were not encoded correctly. As a result, this participant was excluded from the MRI analysis, leaving the final fMRI sample size at 35. Sample size determination was guided by both behavioral and neuroimaging power considerations. For behavioral effects, power analysis using G*Power (Faul et al., 2009) based on preliminary data from the pilot study indicated that 73 participants would provide >80% power to detect medium-sized effects ($d > 0.4$) at $\alpha = 0.05$. Our final behavioral sample of 89 participants exceeded this requirement. For the fMRI component, our sample size of 35 participants was determined based on successful detection of music-modulated sequence learning effects in our previous study (Ren and Brown, under review) using a similar sample size. Prior to the study, all participants completed a self-report questionnaire to ensure they did not have any of the following conditions that could affect the study outcome: auditory impairments, visual impairments not corrected by glasses or contacts, fundamental music processing deficits (e.g., amusia or music agnosia), learning disabilities, attention disorders, or a history of neurological or psychiatric illnesses. Participants who volunteered for the fMRI scan were subject to further screening to exclude any magnetic resonance imaging contraindications

such as metallic implants or major neurological disorders, following the Georgia Institute of Technology/Georgia State University Center for Advanced Brain Imaging's scanning policy. All participants provided informed consent to the procedures approved by the Institutional Review Board of Georgia Institute of Technology.

2.2. Materials

2.2.1. Visual Stimuli

Visual stimuli comprised real-life images selected from diverse databases, including the 360-color items dataset (Moreno-Martínez & Montoro, 2012), MIT indoors scenes (Quattoni & Torralba, 2009), and visual stimuli used and shared by the lab in prior studies. These images encompassed categories such as objects, animals, and scenes with the goal that memory results would better reflect what one might expect for the types of complex and often arbitrary associations experienced in the course of daily life. Each image was adjusted to a size of 300x300 pixels. A total of 72 distinct images were chosen to form 18 sequences, each composed of four images. This sequence length was chosen based on previous research (Ren et al., 2024) indicating that four items per sequence allow for sufficient learning within a two-hour session without reaching a ceiling effect. 12 sequences were assigned to music condition and the remaining 6 to the control condition (see details in Procedure). This allocation was determined by the broader research project investigating different types of music. For the purposes of this dissertation, all music conditions are collapsed into a single 'music present' condition to address our primary research question about the general effects of musical context on sequence learning. The control sequences provide a baseline measure of sequence learning without musical accompaniment. Analyses comparing specific musical features will be discussed in Future Directions in Chapter 4. The object, scene, and animal categories were balanced across the music

and control conditions, as well as across sequences within each condition. However, the specific composition of each sequence varied to avoid introducing additional regularities that could influence learning. For example, while each sequence contained no more than one scene image, some sequences did not include any scenes. The assignment of specific images to sequences was randomized to control for potential biases. An additional set of 50 images, aligned with the sequence categories, served as lures during encoding, interspersed between sequences.

2.2.2. Music Stimuli

Music stimuli were selected from a recently published famous melodies stimulus set (Belfi & Kacirek, 2021) and popular classical music compositions in the literature, for a total of 12 melodies. Based on my prior study (Ren et al., 2024), music familiarity was an important factor for memory outcomes; to select these 12 melodies I first selected the top 20 melodies with highest 'familiarity' score that resonated with both young and old participants from Belfi and Kacirek, 2021. Additionally, I incorporated ten classical pieces frequently used in music literature. I then put these 30 music pieces in to a pilot survey before running the experiment, asking an independent set of 20 people to rate the familiarity of each song. Questions included: 1) rate how fast you can recognize the music from not until the end to immediately; 2) if the music is paused at the fifth second, can you retrieve the melody and keep singing the rest of the song; 3) can you sing along the music when you listen to it either out loud or in your mind? 4) rate the overall familiarity from 0 indicating you don't know the song to 10 indicating you can recall and sing this song out loud without any cue. Following this validation survey with 20 participants, I selected the 12 melodies with the highest familiarity score to be used in this study (participants in the experiment sample further validated their familiarity to be eligible; 2.3.1):

1. "Blue Danube"

2. "Deck the Halls"
3. "Eine Kleine"
4. "Für Elise"
5. "Harry Potter Theme Music"
6. "Itsy Bitsy Spider"
7. "Jingle Bells"
8. "Ode to Joy"
9. "Alla Turca" from Sonata No.11 in A major, K331 (Mozart)
10. "Pop Goes the Weasel"
11. "Rudolph the Red-Nosed Reindeer"
12. "Wedding March" from 'Lohengrin'

These 12 melodies were paired with 12 visual sequences. To control other relevant variables such as volume, timbre, and duration, I employed music composition software (Logic Pro) to extract the core melody of these songs, which were played on the piano. The melodies were then standardized to a volume of 90 dB, with harmonies set to 70 dB for all compositions. I adjusted the tempo of each melody to maintain a uniform duration of 8 seconds for all music stimuli. Each melody was arranged into four distinct musical phrases of equal duration (2 seconds each), with a strong metrical accent at the beginning of each phrase to align with the onset of each image in the sequence. This metric structure ensured consistent temporal alignment between musical phrases and visual stimuli presentation.

For the lure images (non-member images of the sequences) inserted during the encoding of visual image streams, I used Mozart's piano sonata No. 16 as background sound. This piece was chosen as it met several practical criteria: it has sufficient length for the lure presentation

duration (144 seconds), maintains consistent musical style and instrumentation with the sequence-paired stimuli, and was not among the highly familiar melodies used in the main task to avoid confusion. This music was generated using Logic-Pro in the same way as other music stimuli, except for a longer duration (144 seconds).

2.3. Procedure

The experiment consisted of four main components: 1) pre-task screening (> 3 days before the task); 2) practice trials (5 minutes); 3) encoding phase with embedded retrieval practice (in scanner, around 2 hours); 4) final retrieval task (post-scan, 10 minutes)

2.3.1. Participant Pre-Task Screening

Prior to participating, all participants underwent an online pre-task screening questionnaire using Google Form at least three days in advance. They were required to listen to audio clips of the 12 pre-selected music stimuli and rate their familiarity on a scale of 1 to 5 with the questions "Are you familiar with the music?" and “Can you sing along with the melody without any problem?”. Only participants who answered all question with a score of 5 were eligible to participate in the task.

2.3.2. Experimental Procedure

The entire experiment was conducted using MATLAB Psychtoolbox 3 and comprised two distinct phases: encoding and retrieval. The encoding phase took place within the MRI scanner, where participants were exposed to continuous streams of images, some of which were accompanied by music. Participants were instructed to learn the sequential relationships among the images and to indicate their perception of event boundaries. Retrieval practice tasks were inserted within the encoding phase, and a subsequent sequence recall task was used after the

scanning session. These retrieval tasks assessed participants' memory for the learned sequences and individual items. The combination of encoding and retrieval phases allowed for a comprehensive examination of the effects of musical context on temporal order learning and memory.

2.3.2.1. Encoding Phase

The encoding phase took place within the MRI scanner. Participants were instructed that they would be presented with a continuous stream of images and their goal was to extract any regular patterns and learn the sequential relationship of images. To demonstrate their statistical learning of these patterns, they were instructed to perform a segmentation task during image presentation, pressing a button on an MRI-compatible response box whenever they sensed a boundary or breaking point of an event.

The encoding phase contained 18 repeating sequences embedded in a continuous stream of images, surrounded by lure stimuli that occurred stochastically. Among these, 12 sequences were paired with music, with each sequence consistently corresponding to the same music composition for one participant. The sequence-music pairings were randomly assigned at the start for each participant and remained constant throughout. The remaining six sequences were learned without any accompanying music, serving as a control condition. While my previous studies have used an active non-musical control (monotonic tone sequences) to better match engagement of auditory cortex across conditions (Ren et al., 2024), the current study opted for a more naturalistic silence condition. This choice was made to avoid potential confounds introduced by rhythmic stimuli, which could still aid learning more than silence does, and to better represent real-world learning scenarios where visual sequences are often encountered without any auditory accompaniment.

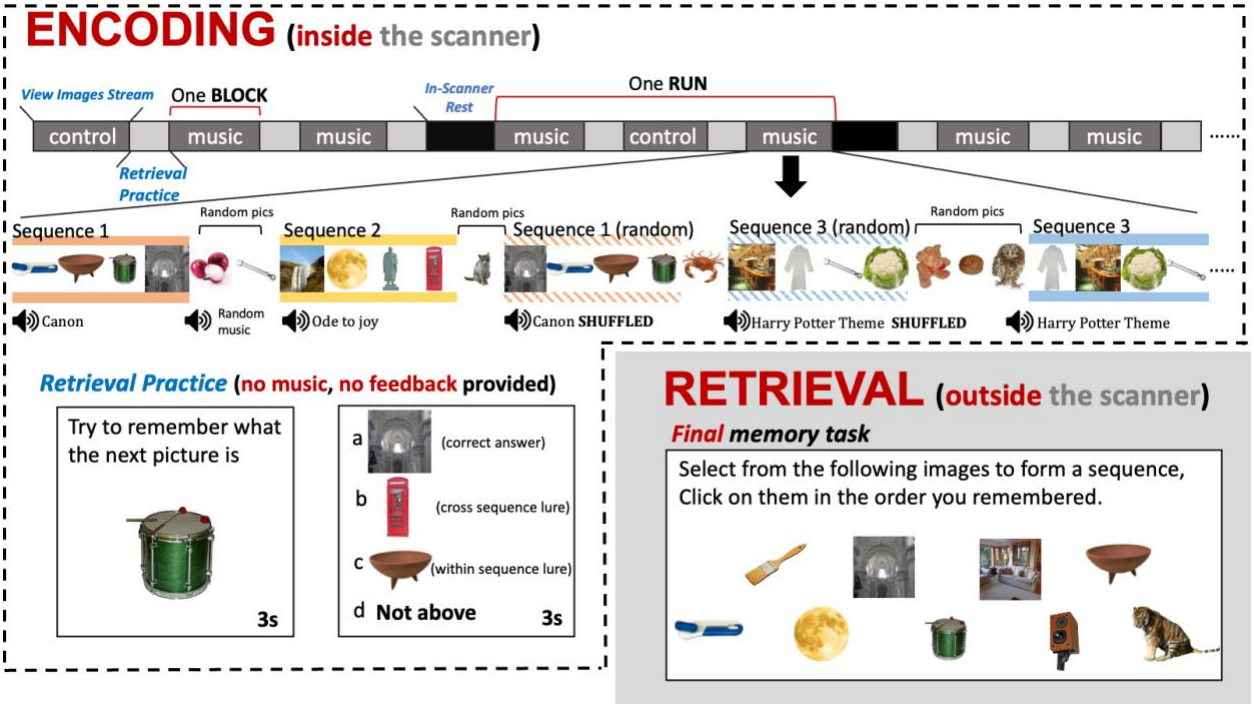

Figure 1 Experiment Procedure

During encoding, participants viewed continuous streams of images paired with music in some blocks and not in others. Sequences were probabilistically ordered (2/3 correct and 1/3 shuffled). Music played in sync with each image for the music condition. Participants were asked to segment events by pressing a button when they perceived a boundary. Retrieval tasks followed each block, testing sequence order recall. After fMRI scanning, a retrieval task was given where participants selected from provided images and ordered them into a sequence they learned.

During encoding, shown in **Figure 1**, participants viewed continuous streams of images. Each image was presented for 2 seconds, and each sequence comprised four images, resulting in a total duration of 8 seconds per sequence. There was a fixed 2/3 probability that each sequence would be displayed in the correct fixed order, while there was a 1/3 probability that the sequence would be randomly shuffled. This specific probability was chosen based on a combination of pilot testing and practical considerations. Pilot studies were conducted with varying probabilities

to identify a level of difficulty that was challenging enough to require statistical learning but not so difficult as to make learning impossible within the given time constraints. Moreover, the 2/3 probability allowed for a balanced design that could accommodate the desired number of trials within the 2-hour scanning session. In the presence of music, each musical excerpt was also 8 seconds long, corresponding to the duration of a single sequence. If the sequence was shuffled, the music was also correspondingly shuffled—each two seconds of the music paired with a specific picture within the sequence. To prevent participants from simply segmenting events every eight seconds, various distracting images were interspersed between sequences. Concurrently, a random section from Mozart's piano sonata No.16 was played. This probabilistic presentation of sequences, along with the inclusion of lure images, was designed to create a statistical learning task where participants had to extract regularities from a noisy input stream. By contrast, a deterministic repetition of stimuli would not require the same level of statistical computation and would instead rely more on rote memorization. The current design aims to capture the complex and probabilistic nature of real-world learning scenarios, where perfect repetitions are rare, and learners must discern patterns amidst noise and variability.

The experiment was conducted over six scanning runs, interspersed with in-scanner rest periods and scanner restarts. Each RUN lasted approximately eleven minutes. This allowed participants to relax and rest which is critical for their performance due to the length of the scan (2 hours in total). Within each run, participants had three learning BLOCKs, divided by retrieval practice tasks. Among these three blocks, two were music condition and one was control condition. In each block, participants learned three sequences in total, with each sequence appearing six times and having a 1/3 chance of being shuffled. The order of the sequences and

blocks was randomized. Participants completed one repetition of learning all 18 sequences in every two scanning runs, making it three repetitions of learning in total during the scan.

Below summarized the main sequence presentation parameters :

- Probability of correct order: 2/3
- Probability of shuffled order: 1/3
- Number of repetitions per sequence: 6 times per block
- Number of blocks per scanning run: 3
- Number of scanning runs: 6
- Total repetitions of sequences learning across experiment: 3 (across all runs)

2.3.2.2. Retrieval Practice Tasks

Retrieval practice tasks were inserted after every three and half minutes of image presentation (after one block where three sequences were learned), resulting in three retrieval practice rounds per run. The inclusion of these tasks served two crucial purposes. First, based on early pilot studies, it was found that participants experienced difficulty maintaining focus and learning if the continuous stream of images was too long. Therefore, the decision was made to divide each run into three blocks, with retrieval practice tasks providing a mental break for participants. More importantly, while segmentation presses offered insight into participants' learning in terms of their ability to segment and cluster visual items, the retrieval practice tasks provided a measure of their learning of the temporal order relationships between items. This additional information was essential to gauge participants' progress in extracting temporal patterns, a key aspect of statistical learning that goes beyond mere segmentation. During retrieval practice, participants were presented with a picture from the previous image stream and were prompted to recall the subsequent image. The cue remained on screen for three seconds.

Participants then saw four options: the accurate image, a luring image from the same sequence (within-sequence lure), a luring image from a different sequence within the same prior image presentation (cross-sequence lure), and a fourth option "none above" (correct when the cue image was the final image of a sequence). Participants were asked to make a choice within 3 seconds. Each retrieval practice entailed six questions, with the cue image randomly chosen from previously viewed image streams. No music was involved in this stage.

2.3.2.3. Final Retrieval Task

Following the encoding phase, participants exited the scanner and engaged in two memory retrieval tasks on a computer, each took around 3-5 minutes. The first was an item recognition task. Thirty items randomly selected from the encoding task and additional twenty new images were used. For each trial participants would see an image and they needed to decide if they have seen that image during the task. One purpose of this task was to build a time buffer to enable a test of delayed recall of the sequence memory (next task), and to provide an assessment of individual item familiarity which could help interpret performance levels in that more challenging recall task (i.e., if they don't recognize the individual items well we can expect poor performance recalling the relational information). Then they performance a sequence retrieval task. Each of the 18 sequences from encoding was tested once. For each trial, participants viewed nine images on-screen for each question. Among these, four belonged to one sequence, while the remaining five were randomly selected from other sequences learned within the same run as the to-be-tested sequence. Participants had to select, group, and reorder these images into a familiar sequence by clicking on them sequentially.

Notably, no music or feedback was provided during retrieval practice or the final retrieval task, restricting interpretation of measured impacts of the music manipulations on memory to the encoding phase.

2.3.2.4 Practice Trial

Prior to the main encoding phase, participants completed a 5-minute practice trial to familiarize themselves with the task structure. The practice consisted of four sequences presented in a format identical to the main task: each sequence appeared six times with a 1/3 probability of random order shuffling. The practice sequences were presented without musical accompaniment to establish baseline task understanding. Following the practice sequences, participants completed a sequence reconstruction task where they grouped and ordered the images they had observed. This practice phase served two purposes: (1) to familiarize participants with the continuous stream presentation and segmentation response requirements, and (2) to demonstrate the task goal of identifying and learning sequential relationships among images. All practice stimuli were selected from the same image database as the main task but were not used in the experimental trials to prevent interference effects.

2.4. fMRI acquisition

Brain scans were collected using a 3T Siemens Prisma scanner with a 32-channel head coil at the GSU/GT Center for Advanced Brain Imaging. High-resolution T1-weighted structural images were obtained using magnetization prepared rapid gradient echo (MPRAGE) sequences with a repetition time (TR) of 2500 ms, echo time (TE) of 2.22 ms, and field of view (FOV) of 256 × 240 mm. The sequence used 0.8 mm slice thickness with 0.8 mm isotropic voxel size, an 8° flip angle, and 220 Hz/pixel bandwidth. GRAPPA parallel imaging was employed with an acceleration factor of 2.

T2-weighted structural images were acquired using 3D SPACE sequence with GRAPPA parallel imaging (acceleration factor = 2). The sequence parameters included a TR of 3200 ms, TE of 563 ms, and FOV of 256 × 240 mm. The images were acquired with 0.8 mm slice thickness, 0.8 mm isotropic voxel size, 120° flip angle, and 744 Hz/pixel bandwidth.

T2*-weighted functional data and localizer data were obtained using Multiband echo planar imaging (EPI) sequences with an acceleration factor of 5. The sequence employed a TR of 750 ms, TE of 32 ms, and FOV of 220 × 220 mm. Images were acquired with 2.5 mm isotropic voxel size across 72 slices, using a 52° flip angle and 2367 Hz/pixel bandwidth. The T2*-weighted images provided whole-brain coverage and were acquired with slices oriented parallel to the long axis of the hippocampus.

Detailed hippocampal scans were obtained in T2 space with selective excitation using ZOOMit 2D TSE (Turbo Spin Echo) provided by Siemens. The sequence parameters included a TR of 4560 ms, TE of 51 ms, 2 mm slice thickness, 130° flip angle, and 100 Hz/pixel bandwidth.

All participants were provided with foam ear plugs and a noise-cancelling MRI-compatible headset. This equipment served dual purposes: enabling communication with the experimenter between runs in the scanner and optimizing music stimuli delivery during scanning. Although all music stimuli were standardized to the same volume level, participants were asked to adjust the sound volume inside the scanner using an MRI-compatible controller before task initiation. They were instructed to set the volume to a level where the music was clearly audible in the MR environment while maintaining comfort and safety. Consequently, while the absolute intensity level of the sound varied across participants, the relative intensity remained consistent across all audio stimuli within each participant's session.

2.5. fMRI preprocessing

Results included in this manuscript come from preprocessing performed using fMRIPrep 23.2.1 (Esteban et al., 2019) (RRID:SCR_016216), which is based on Nipype 1.8.6 (Gorgolewski et al., 2011) (RRID:SCR_002502).

2.5.1. Preprocessing of B0 inhomogeneity mappings

A total of 1 fieldmaps were found available within the input BIDS structure for this particular subject. A B0-nonuniformity map (or fieldmap) was estimated based on two (or more) echo-planar imaging (EPI) references with topup (Andersson et al., 2003).

2.5.2. Anatomical data preprocessing

A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1w image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.5.0 (Avants et al., 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL (version unknown), RRID:SCR_002823, Y. Zhang et al., 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 7.3.2, RRID:SCR_001847, Dale, Fischl, and Sereno 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al., 2017). A T2-weighted image was used to improve pial surface refinement. Brain surfaces were reconstructed using recon-all (FreeSurfer

7.3.2, RRID:SCR_001847, Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al. 2017). Volume-based spatial normalization to two standard spaces (MNI152NLin6Asym, MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.5.0), using brain-extracted versions of both T1w reference and the T1w template. The following templates were were selected for spatial normalization and accessed with TemplateFlow (23.1.0, Ciric et al. 2022): FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model [Evans et al. (2012), RRID:SCR_002823; TemplateFlow ID: MNI152NLin6Asym], ICBM 152 Nonlinear Asymmetrical template version 2009c [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

2.5.3. Functional data preprocessing

For each of the 4 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume was generated, using a custom methodology of fMRIPrep, for use in head motion correction. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL , Jenkinson et al. 2002). The estimated fieldmap was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with six degrees of freedom. Several confounding

time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al. (2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's aseg segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the

correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance timeseries are calculated by means of principal components analysis of the signal found within a thin band (crown) of voxels around the edge of the brain, as proposed by (Patriat et al., 2017). All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using nitransforms, configured with cubic B-spline interpolation.

Copyright Waiver

The above boilerplate text in this [fMRI processing] section was automatically generated by fMRIPrep with the express intention that users should copy and paste this text into their manuscripts unchanged. It is released under the CC0 license.

2.6. fMRI analysis

The fMRI analysis employed multiple complementary approaches to characterize how music influences sequence learning at both regional and network levels. We began with traditional univariate analyses using generalized linear models (GLM) to identify brain regions showing differential activation across task conditions. To capture temporal dynamics, particularly at sequence boundaries, we implemented finite impulse response (FIR) models. Given our specific hypotheses about the roles of memory-related regions, we conducted focused analyses within anatomically-defined regions of interest (ROIs). To understand how these

regions work together, we examined functional connectivity patterns during learning. Finally, we used representational similarity analysis (RSA) to investigate how neural patterns within the hippocampus reflect the structure of learned sequences and how these representations are modulated by musical context.

2.6.1. Univariate Generalized Linear Model

After preprocessing, a generalized linear regression model (GLM) analysis was performed to understand how functional BOLD responses were modulated by different task conditions. The GLM analysis utilized FSL (FMRIB Software Library, Version 6.0.5.6; Jenkinson et al., 2012), specifically FEAT (FMRI Expert Analysis Tool) for first- and higher-level analyses (Woolrich et al., 2004).

Before running the GLMs, the preprocessed data from fmrirep was further smoothed with a 5mm full-width at half maximum Gaussian kernel in FEAT. The model included six nuisance regressors accounting for head motion (translational and rotational), twelve possible image/task conditions during the visual stream presentation (context: music or control; sequence presentation order: ordered or shuffled; position of the image: within-sequence (image 1-3) or boundary (image 4)), and four conditions during retrieval practice (See ***Table 1***). The task regressors were convolved with a double-gamma hemodynamic response function, and temporal derivatives were included. The complete design matrix comprised 38 regressors per run.

Group-level analyses were performed using FSL's FLAME (FMRIB's Local Analysis of Mixed Effects) stage 1 to account for both within-subject and between-subject variability. Final cluster results were thresholded using a cluster-corrected threshold of $Z > 3.1$, with a cluster significance threshold of $p < 0.05$.

Table 1 Regressor Labels and Conditions for GLM

Regressor Label	Sequence Learning		
	Context Condition	**Sequence Order**	**Location of the Image**
CO_WT	Silence/Control	Ordered	Within Event
CO_BD	Silence/Control	Ordered	Boundary
CS_WT	Silence/Control	Shuffled	Within Event
CS_BD	Silence/Control	Shuffled	Boundary
MO_WT	Music	Ordered	Within Event
MO_BD	Music	Ordered	Boundary
MS_WT	Music	Shuffled	Within Event
MS_BD	Music	Shuffled	Boundary
Lure_CO	Silence/Control	Ordered	Non-member Lure
Lure_CS	Silence/Control	Shuffled	Non-member Lure
Lure_MO	Music	Ordered	Non-member Lure
Lure_MS	Music	Shuffled	Non-member Lure
	Retrieval Practice		

	Image Condition during Learning	Task Phase
Cue_C	Control	Cue shown
Ans_C	Control	Subjects make answer
Cue_M	Music	Cue shown
Ans_M	Music	Subjects make answer

2.6.2. Regions of Interest

Based on the literature review above and prior studies in our lab, I focused on six regions of interest (shown in **Figure 7** right bottom corner): hippocampus, parahippocampal gyrus, ventromedial prefrontal cortex (vmPFC), inferior frontal gyrus (IFG), and the striatum (specifically NAc and caudate). The IFG, parahippocampal gyrus, caudate, and NAc masks were extracted in FSL using the Harvard-Oxford Atlas (Desikan et al., 2006). The hippocampus masks were manually traced for each participant using ITK-SNAP (www.itksnap.org; Yushkevich et al., 2006). Our team meticulously delineated the hippocampus body, including subfields CA1, CA2, CA3, dentate gyrus, and subiculum, for both the right and left hippocampus on high-resolution hippocampus scans using the finalized protocol developed by the Hippocampal Subfields Group (*in prep*; https://hippocampalsubfields.com/harmonized-protocol/) (**Figure 2**.a). To align these traced masks with other neuroimaging data, I employed a two-step registration process. First, FSL's FLIRT (FMRIB's Linear Image Registration Tool) was used to calculate the transformation matrix between the high-resolution hippocampus scan and the T1 scan, allowing us to realign the traced masks into the T1 native space. Subsequently, these

realigned masks were transformed into MNI (Montreal Neurological Institute) standard space using ANTs (Advanced Normalization Tools), ensuring spatial consistency across participants for group-level analyses.

For the vmPFC, considering its broad anatomical coverage, the masks were 5mm spheres centered on coordinates reported by the lab's prior studies (anterior vmPFC: MNI = [-2, 56,-10]; posterior vmPFC: MNI = [-5, 16, -11], reported in Brown et al., 2016), which our ongoing work implicates in using schemas to support new learning.

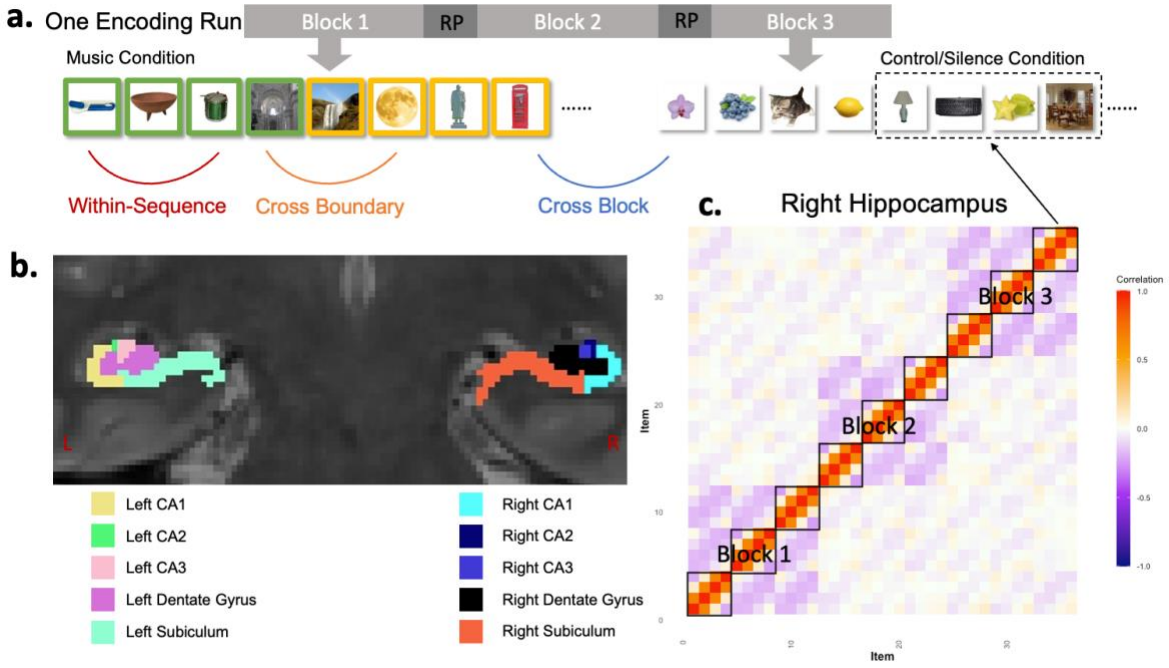

Figure 2 Representational Similarity Analysis

(a) *Pattern Similarity for Different Item Relationships in Hippocampal Subfields.* Three types of item relationships were used in the representational similarity analysis (RSA): 1) items within the same sequence, 2) cross-sequence items within the same block, and 3) cross-block items. The pattern similarity for these relationships was computed for each hippocampal subregion, allowing investigation into how hippocampal subfields process sequential and cross-boundary relationships

under different conditions (music vs. control). **(b)** *Hippocampal Subfield Comparison for Music vs. Control Conditions*. Subfields were traced on high resolution hippocampus scan using ITK-SNAP. (**c)** *Averaged Representational Similarity Matrix (RSM) for Right Hippocampus.* This matrix presents the averaged representational similarity results for the 36 items in the right hippocampus body. Items learned with music (#1-24) showed positive correlation for within-sequence items and negative correlation for cross-sequence items learned within the same block.

2.6.3. Finite Impulses Response Analysis

To characterize the temporal dynamics of neural responses during sequence boundary processing, we employed finite impulse response (FIR) models. While fMRI's temporal resolution is limited, FIR models offer an advantage over standard GLM analyses by examining the timing of neural responses without assuming a specific hemodynamic response shape. This approach was particularly valuable for investigating how boundary-related neural patterns evolved with sequence familiarity and musical context.

The FIR model parameters were selected to optimize temporal resolution while maintaining adequate signal-to-noise ratio. We used 1-second time bins across an 8-second duration window following the onset of each sequence's final image. The model included four distinct conditions: control condition with ordered sequences, control condition with shuffled sequences, music condition with ordered sequences, and music condition with shuffled sequences. Each condition was modeled using eight FIR regressors, corresponding to the eight 1-second time bins in the analysis window.

The analysis was implemented in two stages. First-level analysis was performed for each run using FSL FEAT, providing individual subject-level analysis of temporal dynamics. Second-level analysis was conducted separately for each learning stage (runs 1, 2, and 3), enabling

examination of how boundary-related neural patterns evolved across the learning period. This separation was crucial for understanding the progression of learning effects over time.

Based on our hypotheses regarding boundary processing, we focused our analysis on three key regions: the medial temporal lobe (MTL), the inferior frontal gyrus (IFG), and the ventromedial prefrontal cortex (vmPFC). These regions were selected based on their established roles in hierarchical sequence learning and boundary-related processing from previous literature (Ezzyat & Davachi, 2021; Schapiro et al., 2013). Two participants were excluded from this analysis because partial runs were completed outside the scanner, resulting in a final sample size of 33 participants for the FIR analyses.

2.6.4. Functional Connectivity

One aim of this study was to understand how different brain regions interactively contribute to statistical learning and sequential learning, and how the networks were modulated by music. To address this question, I performed a seed-based functional connectivity analysis and an ROI-to-ROI connectivity analysis (including graph theory) using CONN Toolbox (Whitfield-Gabrieli & Nieto-Castanon, 2012). Data preprocessing included CONN's default denoising pipeline to minimize the influence of physiological and motion-related confounds. This pipeline implemented the anatomical CompCor approach (Behzadi et al., 2007), which included: (1) regression of white matter and CSF signals using principal component analysis (5 components each), (2) regression of motion parameters and their first-order temporal derivatives, (3) identification and scrubbing of motion-affected time points (frame-wise displacement > 0.5 mm), (4) linear detrending, and (5) band-pass filtering (0.008-0.09 Hz). For seed-based gPPI analysis, I limited the ROIs to: hippocampus, amygdala, parahippocampal gyrus, caudate, NAc and vmPFC. For the ROI-to-ROI analysis, I limited ROIs to: hippocampus, amygdala,

parahippocampal gyrus, inferior frontal gyrus, caudate, putamen, NAc, superior frontal gyrus, medial frontal gyrus, angular gyrus, and vmPFC. The vmPFC masks were again derived from prior schema-related studies, using 5mm spheres. Other ROIs were created directly using CONN's default AAL atlas (Tzourio-Mazoyer et al., 2002). For functional connectivity analysis, I mainly focused on the contrast between control and music during the whole encoding period because this comparison would reveal how music systematically altered the coordination between memory encoding system and other distinct networks such reward systems, during sequence learning, potentially explaining music's facilitatory effects on memory formation.

2.6.5. Representational Similarity Analysis (RSA)

RSA has been used in statistical learning and sequential learning tasks because it can provide further evidence of how memories of items are associated and distinguished. In this task, I had a particular hypothesis that music could provide stronger context to associate events within a sequence, as well as provide clearer boundaries to distinguish sequences. Thus, I used RSA to test whether neural pattern similarity between items exhibited evidence for these groupings and boundary definitions when comparing representational similarity for items with different sequential distances during the learning period and for items from different conditions. In this analysis, I was particularly interested in the hippocampus and its subregions given their established roles in pattern separation, sequence learning and association learning (Howard & Eichenbaum, 2015; Leutgeb et al., 2007; Rolls, 2013). For example, dentate gyrus (DG) subfield is theorized to perform pattern separation to minimize interference between similar memories (Berron et al., 2016; Leutgeb et al., 2007). CA1, with its role in temporal order encoding and context processing(Barrientos & Tiznado, 2016; Hoge & Kesner, 2007), might be especially sensitive to musically-defined temporal context .Therefore, RSA of hippocampal subfields could

reveal whether and how musical context influences both local (within-sequence) and global (between-sequence) organizational principles of memory representations.

Previous studies have compared systematic biases in similarity that can arise from different approaches to modeling single-trial parameter estimates (Mumford, 2013); inspired by the "Least-squares single" approach which mitigates these biases, I decided to run separate GLMs for each of the 72 items within each run during the scan for each participant using FSL's FEAT. In this framework, for each GLM, one single item/picture was modeled as the regressor of interest, and the rest of the pictures formed another regressor. Six nuisance regressors of motion were also included. Voxel-wise betas were then extracted for each item (from each model) for the ROI(s). Unsmoothed preprocessed data from fmriprep was used to preserve fine-grained patterns and spatial specificity to the ROI in the data. Due to some participants' excessive motion during high resolution hippocampus scan, sample size of this analysis is 30.

RSA was conducted in R to compare the voxel-wise patterns of neural activity corresponding to different items in the ROIs. For each participant, within each ROI, a representational similarity matrix (RSM) was computed by correlating the voxel-wise beta patterns across all 72 items per run using Pearson's correlation. The items were categorized into two learning conditions (music and control). Three item relationships were focused, shown in **Figure 2**.a, I specifically extracted within-sequence items pattern similarity, cross-boundary item pattern similarity and cross-block item pattern similarity for each item and averaged each type to obtain a single number for each item for each type. Repeated Measure ANOVA was used to test how context condition and item relationship type together affected pattern similarity for each hippocampal subregion (ten in total, shown in **Figure 2**.b). Tukey's HSD pairwise test was used post-hoc.

To investigate how hippocampal pattern similarity related to memory performance, I examined the relationship between neural patterns during encoding and subsequent retrieval practice performance. For each tested item during retrieval practice, I extracted its pattern similarity with the subsequent item in the sequence from the encoding-phase RSMs. These similarity values were analyzed using a mixed-effects model implemented with the lme4 package in R. The model included fixed effects for retrieval accuracy (forgotten/remembered), context condition (music/control), and encoding run (early/middle/late), with random intercepts for subjects. Pairwise comparisons were conducted to explore condition-specific effects on pattern similarity between remembered versus forgotten associations.

Lastly, I examined whether music enhances position-specific neural representations by conducting a binary position classification analyses. For each adjacent position pair (1→2, 2→3, 3→4), I trained a support vector machine (SVM) classifier to distinguish between hippocampal activity patterns associated with each position. Feature selection was performed prior to classification by selecting the top 30% most position-selective voxels based on F-statistics from training data (this number is chosen from multiple attempts). Classification was performed using leave-one-sequence-out cross-validation, where patterns from all but one sequence were used for training, and the held-out sequence was used for testing.

CHAPTER 3. RESULTS

3.1. Behavioral Results

As part of our investigation into whether musical context facilitates temporal event segmentation, participants completed an event boundary detection task during image viewing. During the viewing of images, participants were tasked with pressing a button whenever they perceived an image as the 'end of an event' or an 'event boundary'. It's important to note that participants were not informed about any association between the images and music, and they were explicitly instructed to focus solely on the images and their sequential relationships. Nonetheless, it was hypothesized that participants might benefit from the presence of the music backdrop in segmenting the images into groups. To assess this, the successful event boundary recognition rate was calculated at each event boundary. A successful detection was counted when a participant pressed a button(s) at the last image of the sequence or the luring image(s) following the last image but not pressing buttons during the middle of a sequence. Average detection accuracy was calculated for each participant for each run and each condition (context x presentation order). **Figure 3**.a displays the event boundary detection rates across three encoding runs. A repeated measures ANOVA was conducted testing run, context and presentation order as main effects showed significant effects of all three variables (run: $F(2, 1007) = 28.674$, $p < .001$, $\eta p^2 = .05$; context: $F(1, 1007) = 18.742$, $p < .001$, $\eta p^2 = .02$; presentation order: $F(1, 1007) = 19.038$, $p < .001$, $\eta p^2 = .02$, see details in Appendix ***Table 8***). Pairwise comparison using Bonferroni correction for multiple comparisons revealed consistently higher boundary detection rate in music condition (music ordered > control ordered; music shuffled > control shuffled; see Appendix ***Table 9*** for detailed statistical significance).

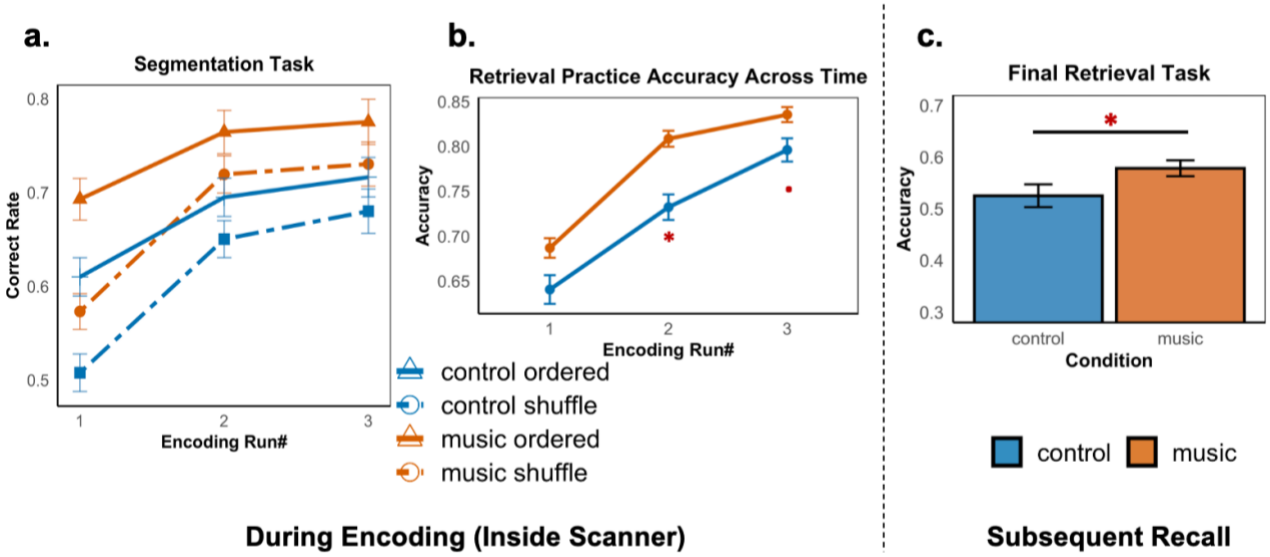

Figure 3 Memory Behaviors

(4) *Event Boundary Detection*: shows the detection rates of event boundaries across three encoding runs for music and control conditions. **(b)** *Retrieval Accuracy:* displays retrieval accuracy across three encoding runs for both conditions, comparing music and control. **(c)** *Final Retrieval Accuracy*: Presents the final sequence retrieval accuracy comparing music and control.

The retrieval practice tests inserted during encoding provided an opportunity to assess participants' learning progress over time. We hypothesized that if music's predictable compositions serve as effective temporal organizing contexts, participants would show enhanced sequence learning in the music condition compared to the control condition. By calculating the correct retrieval rate for both the control and music conditions, an average learning curve was constructed (**Figure 3**.b). The learning curve revealed consistently higher retrieval accuracy for the music condition compared to the control condition across encoding. A repeated measure ANOVA treating subject as repeated measure testing run and context condition as main effect

suggested significant effects of these two variables (run: $F(2, 517) = 42.05$, $p < .001$, $\eta p^2 = .14$; context: $F(1, 517) = 9.522$, $p = .002$, $\eta p^2 = .02$; see Appendix ***Table 10***). A post-hoc pairwise comparison using Bonferroni correction indicated a significant higher retrieval accuracy in music condition at run 2 and a trending effect at run 3 (run 2: mean difference = -0.071, 95% CI [-0.133, -0.009], $p = .026$; run 3: mean difference = -0.053, 95% CI [-0.116, 0.010], $p = .099$; see Appendix ***Table 11*** for details).

After encoding, participants performed an item recognition task. Recognition memory performance was analyzed using signal detection theory to account for both hits and false alarms. For sequence items learned in the music condition, participants showed high discrimination sensitivity (d' = 3.95, SD = 0.88), with a hit rate of 96.4% (SD = 9.1%) and a false alarm rate of 4.9% (SD = 10.5%). Similarly robust performance was observed for sequence items from the control condition (d' = 4.07, SD = 0.88; hits = 96.9%, SD = 9.8%; false alarms = 4.9%, SD = 10.5%). Non-sequence lure items that were not part of any temporal structure also showed comparable discrimination (d' = 3.96, SD = 1.00; hits = 95.7%, SD = 10.7%; false alarms = 4.9%, SD = 10.5%).

A repeated measures ANOVA comparing discrimination sensitivity (d') across the three item types revealed no significant differences between conditions ($F(2,222) = 0.092$, $p = .912$), suggesting that the presence of musical context during encoding did not differentially affect subsequent item recognition. The consistently high d' values across all conditions indicate that participants maintained strong item memory regardless of the learning context or temporal structure. Two participants were identified as outliers (exceeding 2 SD from the mean d') and were excluded from subsequent analyses, leaving a total behavioral results sample size of 87.

Last, in the Final Retrieval Task, participants were required to identify images belonging to the same sequence/event but also to correctly order them. While statistical learning is often considered an implicit learning process, the data suggested the possibility of explicit learning of the sequential relationships between images. This was particularly noteworthy given that the sequential relationships were probabilistic rather than deterministic during encoding, with each sequence appearing in its fixed order only 66.6% of the time. For each sequence, participants received credit only if they reconstructed the entire sequence in its correct order (i.e., all images placed in their exact serial positions). Under these scoring criteria, participants successfully reconstructed 57.8% of sequences (SD = 30.6%) in music condition, while those sequences in the control condition were reconstructed with 52.5% correct rate (SD = 31.7%). Consistent with our hypothesis that musical context would enhance temporal sequence learning, a paired-samples t-test revealed significantly higher sequence reconstruction accuracy in the music condition compared to the control condition ($t(86) = -2.292$, $p = .024$, $d = .246$) (**Figure 3**.c).

3.2. Whole-brain Univariate GLM Results

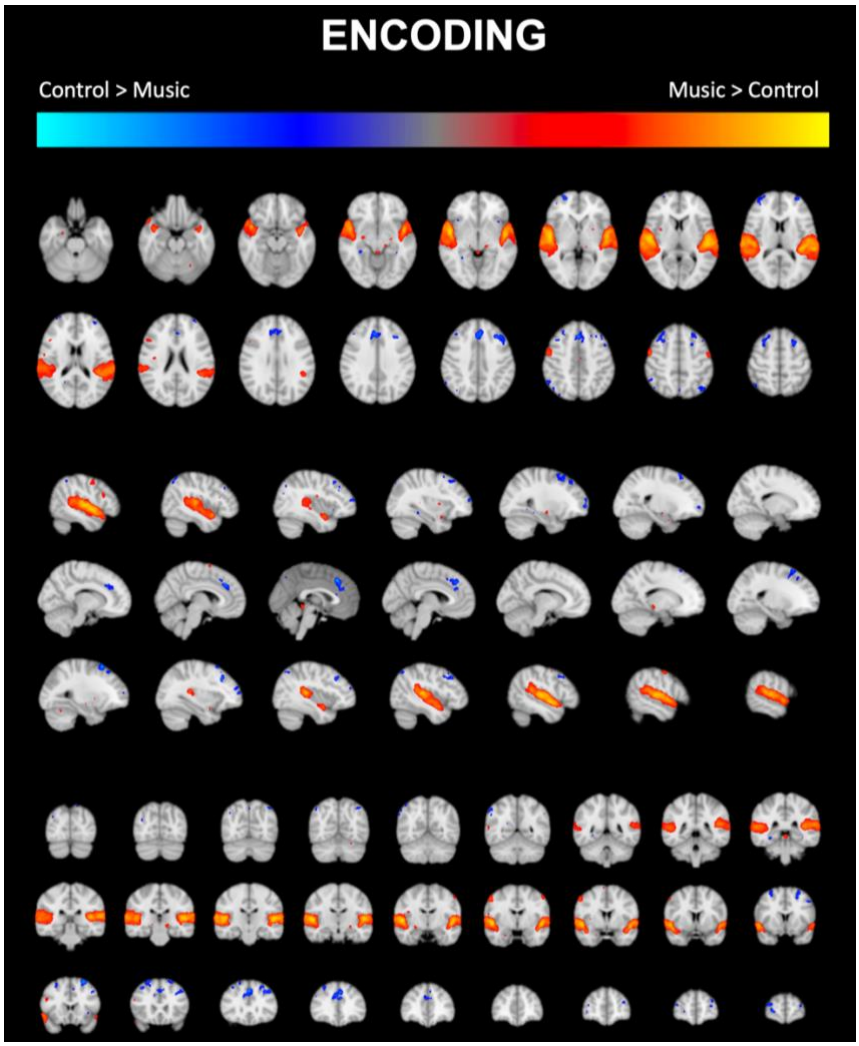

Figure 4 Whole-Brain Univariate GLM Results for Encoding Phase

Regions with significant BOLD activity differences between music and control conditions during the encoding phase are displayed. Control showed stronger activation in prefrontal and occipital regions, while music activated auditory and motor-related areas.

Building on the behavioral findings, to understand the whole-brain neural patterns during the statistical learning and temporal order memory task modulated by music, I conducted

univariate GLM analysis using FSL's FEAT. Based on previous findings suggesting music might reduce neural demands during sequence learning (Ren and Brown, under review), I first examined the contrast between music and control conditions for the entire encoding phase regressors. **Figure 4** shows the regions that exhibited different BOLD activity during the two conditions. Images and sequences learned without music (control) showed stronger BOLD activity in prefrontal regions including the superior frontal gyrus and middle frontal gyrus, as well as the lateral occipital cortex. In contrast, the music condition showed stronger BOLD activity in music and auditory-processing-related regions such as the superior temporal gyrus and precentral gyrus (an expected outcome and useful sanity check, given the control condition in this study was sequencing in silence). Detailed activation clusters are provided in ***Table 2***.

Table 2 Clustering Results for Whole-Brain GLM Contrast during Encoding

Cluster Index	Voxels	P	Z-MAX	Z-MAX X	Z-MAX Y	Z-MAX Z	Brain Region
Control > Music							
1	557	1.10E-09	5.62	2	28	42	Paracingulate Gyrus
2	391	1.79E-07	4.63	24	24	62	Superior Frontal Gyrus [R]
3	351	6.56E-07	4.6	-34	26	36	Middle Frontal Gyrus [L]
4	260	1.62E-05	4.72	-28	18	60	Superior Frontal Gyrus [L]
5	190	0.00025	4.22	44	-62	56	Lateral Occipital Cortex [R]
6	178	0.000413	4.48	28	62	2	Frontal Pole [R]
7	77	0.05	4.08	-44	-68	50	Lateral Occipital Cortex [L]
Music > Control							
1	5546	0	8.86	52	-10	4	Superior Temporal Gyrus [R]
2	4805	3.08E-43	8.42	-48	-6	-6	Superior Temporal Gyrus [L]
3	195	0.000204	5.12	54	-2	50	Precentral Gyrus [R]

I then investigated the control vs. music contrast separately for the regressors specifically modeling the boundary (BD) vs. within the sequence (WT) to better understand how musical context affects different aspects of sequence processing. The activation maps are shown in **Figure 5**, and the detailed cluster results are presented in ***Table 3*** and ***Table 4***. *At the boundary*, the control > music contrast again revealed stronger activity in frontal and occipital regions.

Music, on the other hand, activated not only the expected auditory regions due to the nature of the stimulus, but now there was also a significant difference in the amygdala and thalamus. When examining activation when processing images *within the sequence*, the control condition showed stronger effects in the lingual gyrus and retrosplenial cortex (precuneus) in addition to prefrontal and occipital regions. The music condition continued to show stronger activity in auditory-related areas, including Heschl's gyrus and precentral gyrus (**Figure 5**, bottom).

Table 3 Clustering Results for GLM Contrast at Event Boundaries (Control vs. Music)

Cluster Index	Voxels	P	Z-MAX	Z-MAX X	Z-MAX Y	Z-MAX Z	Brain Regions
Control > Music *at Boundary*							
1	251	2.15E-06	4.83	-30	18	58	Middle Frontal Gyrus [L]
2	136	0.000593	4.55	-40	-76	42	Lateral Occipital Cortex [L]
3	97	0.00561	3.72	-46	14	42	Middle Frontal Gyrus [L]
4	65	0.0443	4.42	54	-56	44	Lateral Occipital Cortex [R]
Music > Control *at Boundary*							
1	6220	0	8.82	-48	-4	-6	Planum Polare [L]
2	6140	0	8.77	52	-10	2	Heschl's Gyrus [R]
3	206	1.69E-05	5.43	54	-2	50	Precentral Gyrus [R]
4	174	8.09E-05	3.99	50	20	22	Inferior Frontal Gyrus [R]
5	86	0.0111	4.56	-22	-16	-10	Amygdala [L]
6	76	0.0212	4.44	10	-16	10	Thalamus [R]

Table 4 Clustering Results for GLM Contrast Within Sequence (Control vs. Music)

Cluster Index	Voxels	P	Z-MAX	Z-MAX X	Z-MAX Y	Z-MAX Z	Brain Region
Control > Music *within Sequence*							
1	888	2.18E-15	4.76	-12	-104	6	Occipital Pole [L]
2	568	3.35E-11	5.21	10	28	34	Paracingulate Gyrus [R]
3	357	5.96E-08	4.51	24	24	60	Superior Frontal Gyrus [R]
4	277	1.31E-06	4.1	16	-92	-8	Occipital Pole [R]
5	176	0.000117	4.44	-28	58	10	Frontal Pole [L]
6	170	0.000156	4.6	34	60	12	Frontal Pole [R]
7	158	0.000282	4.55	-34	26	36	Middle Frontal Gyrus [L]
8	156	0.000311	4.89	-28	-56	-4	Lingual Gyrus [L]
9	141	0.000669	4.79	28	-46	-8	Lingual Gyrus [R]
10	123	0.00174	4.39	20	-58	18	Precuneus [R]
11	89	0.0121	4.36	54	-62	48	Lateral Occipital Cortex [R]
Music > Control *within Sequence*							
1	3163	4.80E-37	8.38	52	-12	4	Heschl's Gyrus [R]
2	2945	2.78E-35	7.82	-50	-22	8	Heschl's Gyrus [L]
3	158	0.000282	4.79	54	-2	46	Precentral Gyrus [R]

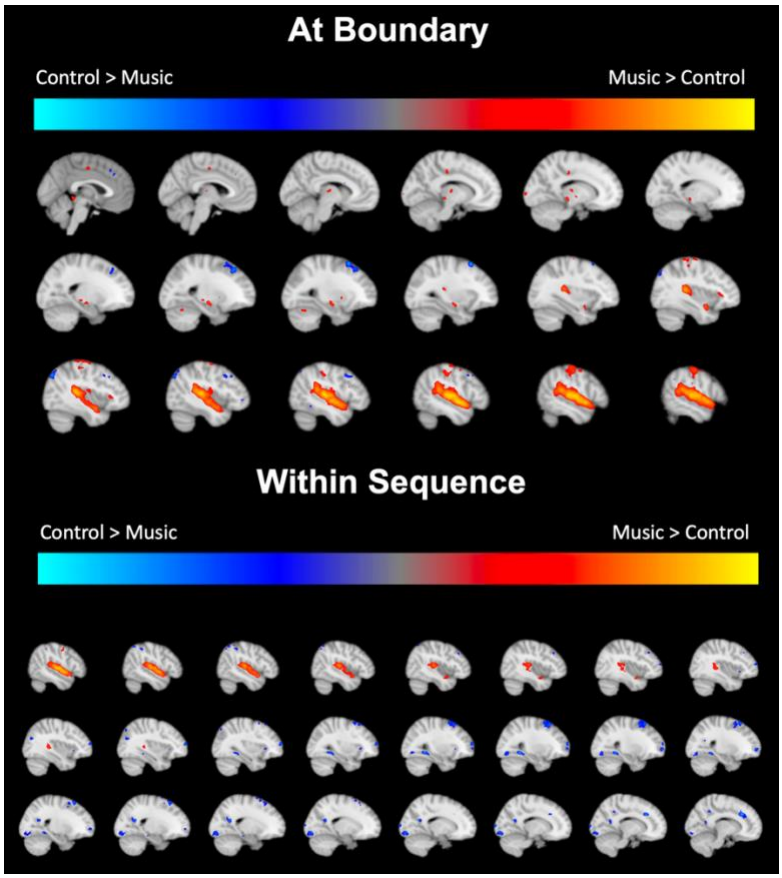

Figure 5 Univariate GLM Contrast – Music vs Control at Event Boundaries and within Sequence

This figure shows the contrast between music and control conditions at sequence boundaries (top) and within sequence (bottom).

Lastly, to investigated how encoding context influences subsequent retrieval processes, I conducted a contrast analysis during the retrieval practice period, focusing on the cue and response phases. As shown in **Figure 6**, during the 3-second cue presentation, trials that were originally learned in silence showed stronger activity in the supramarginal gyrus in both hemispheres compared to those learned with music. During participants made responses for the

retrieval practice task, the control condition exhibited stronger activity in the occipital pole and lateral occipital cortex. The decreased activation in the music condition, particularly in regions associated with visuospatial processing and attention (supramarginal gyrus, occipital cortex), may reflect lower cognitive effort required to retrieve temporal relationships when sequences were initially encoded with the structured temporal framework. This interpretation aligns with our behavioral findings showing enhanced sequence memory in the music condition, suggesting that musical context during encoding may facilitate the formation of more readily accessible temporal representations.

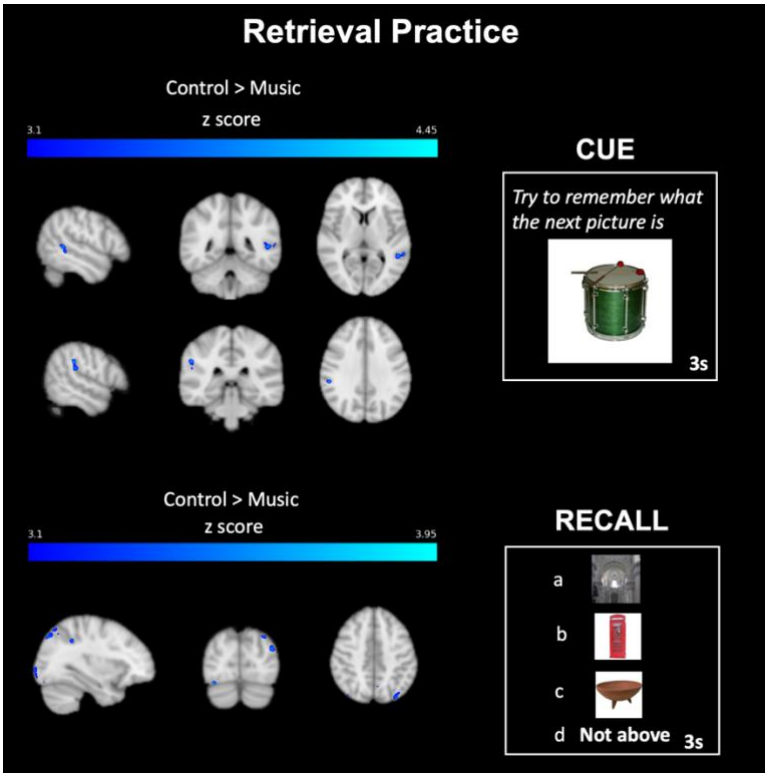

Figure 6 Retrieval Practice Period Univariate Contrast

BOLD activity during retrieval practice is shown for both cue and response phases . The control condition activated the supramarginal gyrus and occipital cortex more strongly, highlighting distinct retrieval strategies between music and control conditions.

In summary, these whole-brain analyses revealed distinct patterns of neural activation between music and control conditions during both encoding and retrieval, demonstrating the substantial influence of musical context on brain activity during statistical learning. The results partially supported our hypotheses: as predicted, we observed reduced neural demands in prefrontal and visual processing regions during music-accompanied learning. However, contrary to our predictions, we did not observe enhanced MTL activity during boundary detection in the music condition. Instead, music uniquely engaged subcortical regions (amygdala, thalamus) during boundary processing. The absence of significant MTL differences between conditions in the whole-brain analysis warrants careful interpretation. Given the MTL's fundamental role in sequence learning and temporal order processing, this region likely remains critically engaged regardless of learning context. This consideration, combined with our specific hypotheses about MTL function in music-accompanied sequence learning, motivated a more focused ROI analysis. The following section examines targeted ROIs – theoretically relevant regions identified from prior literature and our previous work, allowing us to more precisely characterize how musical context modulates activity in these critical memory circuits.

3.3. ROI Analysis Results

3.3.1. Differential BOLD Activity Across Conditions

While the whole-brain analyses provided valuable insights into broad activation patterns, I conducted targeted ROI analyses to test specific hypotheses about how musical context influences key memory circuits. These analyses offered greater statistical sensitivity than whole-brain comparisons by focusing on theoretically motivated regions. The BOLD activity within key regions of interest (ROIs) was extracted from the GLM model and further analyzed to understand how neural responses varied across context (control vs. music), presentation order

(ordered vs. shuffled), and image position (boundary vs. within-sequence). A repeated measures ANOVA was used to address this question, and revealed that these regions were modulated differently by these factors. ***Table 5*** summarizes the ANOVA results.

Table 5 ANOVA Results for ROIs' BOLD Activity Across Conditions

ROI	Context (F, p)	Order (F, p)	Position (F, p)	Context:Order (F, p)	Context:Position (F, p)	Order:Position (F, p)	Cont:Ord:Pos (F, p)
Left Hippocampus	0.054, 0.817	0.206, 0.650	7.638, 0.006**	0.251, 0.617	0.337, 0.562	1.431, 0.233	0.225, 0.636
Right Hippocampus	0.852, 0.357	0.010, 0.919	8.843, 0.003**	0.844, 0.359	2.177, 0.141	3.336, 0.069	0.448, 0.504
IFG	18.627, <.001***	1.431, 0.233	65.624, <.001***	4.651, 0.032*	14.891, <.001***	2.548, 0.112	0.627, 0.429
Anterior VMPFC	5.812, 0.017*	0.034, 0.854	1.721, 0.191	0.552, 0.458	1.241, 0.266	7.808, 0.006**	0.776, 0.379
Posterior VMPFC	11.859, <.001***	0.016, 0.901	0.900, 0.344	0.617, 0.433	4.601, 0.033*	4.439, 0.036*	2.451, 0.119
Anterior PHC	5.286, 0.022*	0.162, 0.688	2.754, 0.098	0.150, 0.699	0.391, 0.532	7.259, 0.008**	0.041, 0.841
Posterior PHC	0.306, 0.580	0.945, 0.332	9.351, 0.002**	0.232, 0.630	0.981, 0.323	0.002, 0.960	0.158, 0.691
NAc	2.356, 0.126	0.734, 0.393	35.632, <.001***	0.838, 0.361	5.727, 0.018*	0.004, 0.949	1.379, 0.242
Caudate	0.145, 0.703	0.036, 0.851	7.790, 0.006**	0.499, 0.481	0.740, 0.390	0.964, 0.327	0.998, 0.319

The hippocampus exhibited a significant main effect of position (left: $F(1, 238) = 7.638$, $p = .006$; right: $F(1, 238) = 8.843$, $p = .003$), with elevated activity for boundary images compared to within-sequence images (**Figure 7**). This boundary-specific enhancement patterns were also observed in posterior parahippocampal cortex (pPHC), nucleus accumbens (NAc), and caudate, suggesting their role in event boundaries processing.

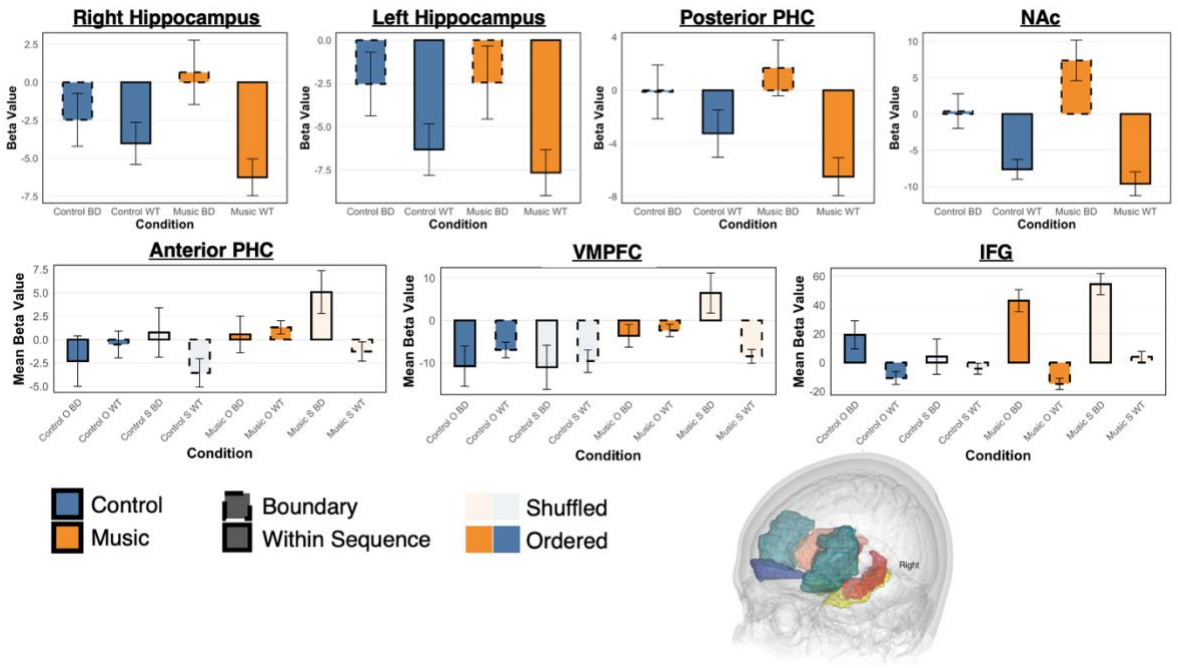

Figure 7 Differential BOLD Activity across Conditions for ROIs

This figure illustrates BOLD activity for regions of interest. It emphasizes the modulation of activity by context (music vs. control), position (within sequence vs. boundary) and presentation order (ordered vs. shuffled).

The ventromedial prefrontal cortex (vmPFC) and anterior PHC showed demonstrated both a main effect of context and an interaction for order and position (aVMPFC: $F(1, 238) = 7.808$, $p = .006$; aPHC: $F(1, 238) = 7.259$, $p = .008$). As shown in **Figure 7**, both regions exhibited higher activity in the music condition than in the control condition. Moreover, both regions showed increased activity for sequence boundaries when presented in shuffled order with music. This pattern suggests these regions may be especially important for integrating temporal context when sequential relationships are less predictable.

Lastly, the inferior frontal gyrus (IFG) was significantly modulated by context ($F(1, 238) = 18.627$, $p < .001$) and image position ($F(1, 238) = 65.624$, $p < .001$), with an interaction effect between these factors ($F(1, 238) = 14.891$, $p < .001$). It showed stronger activity in the music condition compared to the control condition, as well as stronger activity for boundary than within-sequence images, especially in the music context. This aligns with my hypothesis of IFG's role on how musical boundary affects visual boundary. Subsequent sections will continue to investigate this hypothesis.

3.3.2. Correlation Between ROI Activity and Behavioral Performance

To test whether the observed ROI activation patterns reflect mechanisms that directly support sequence learning, I examined the relationship between neural activity and subsequent memory performance. I hypothesized two potential mechanisms: if these ROI effects represent facilitated processing in the presence of music, one would expect negative correlations between BOLD activity and performance, indicating more efficient neural processing. Conversely, if the ROI effects reflect enhanced integration of contextual information during learning, one would expect positive correlations, indicating that greater engagement supports better memory formation. Moreover, these relationships might differ between boundary and within-sequence periods, given their distinct roles in temporal order processing.

I tested these predictions by correlating subject-level beta values and subjects' subsequent retrieval test performance separately for music and control conditions, and for boundary versus within-sequence periods. The results revealed that several regions, including the medial temporal lobe (MTL), vmPFC, and the striatum, exhibited significant associations with behavioral performance during the music condition. In both the hippocampus and the anterior PHC, a positive correlation was observed between BOLD activity and behavioral performance

during the music condition at boundary positions (**Figure 8**, hippocampus (trending): r = .21, p = .085; aPHC: r = .25, p = .036). This suggests that increased activation in MTL at boundaries is associated with enhanced sequence processing. Conversely, both regions showed a negative correlation between BOLD activity and memory performance during the music condition within sequences (hippocampus: r = -.34, p = .004; aPHC: r = -.24, p = .045). These results might suggest the MTL's contribution to memory performance when music is present in the context, but moreover, its neural patterns switched between the boundary and non-boundary events.

The vmPFC showed a significant negative correlation with behavioral performance in the music condition, particularly at within-sequence positions (r = -.43, p < .001). This inverse relationship suggests that lower vmPFC activity during within-sequence image learning is associated with better subsequent memory performance. Similarly, in the striatum, a robust negative correlation was observed (NAc: r = -.45, p < .001; caudate: r = -.27, p = .022) only in the music within-sequence condition, supporting the involvement of this region in encoding the item temporal order relationship. Lastly, in the IFG, a positive correlation was observed only in the control condition at boundary positions (IFG: r = .26, p = .027).

To formally test whether the correlation between context conditions (music vs. control) at different sequence positions (boundary vs. within-sequence) were significantly different, I conducted Fisher's Z-tests. The test revealed that the correlation between the retrieval performance and the hippocampus, posterior PHC, right NAc BOLD activity within sequence were significantly different (more negative) in the music condition (left hippocampus: z = -2.126, p = .033; right hippocampus: z = -2.323, p = .020; posterior PHC: z = -2.165, p =.030; right NAc: z =-2.311, p = .021). A trending difference was found in anterior vmPFC (z = -1.710, p = .087), anterior PHC (z = -1.699, p = .089).

These results support a dual mechanism through which music influences sequence learning: enhanced boundary processing in the MTL (positive correlations) coupled with more efficient within-sequence processing across MTL, prefrontal, and striatal circuits (negative correlations). This pattern aligns with our broader hypothesis that music provides an organizational framework that both strengthens event boundaries and facilitates the encoding of sequential relationships.

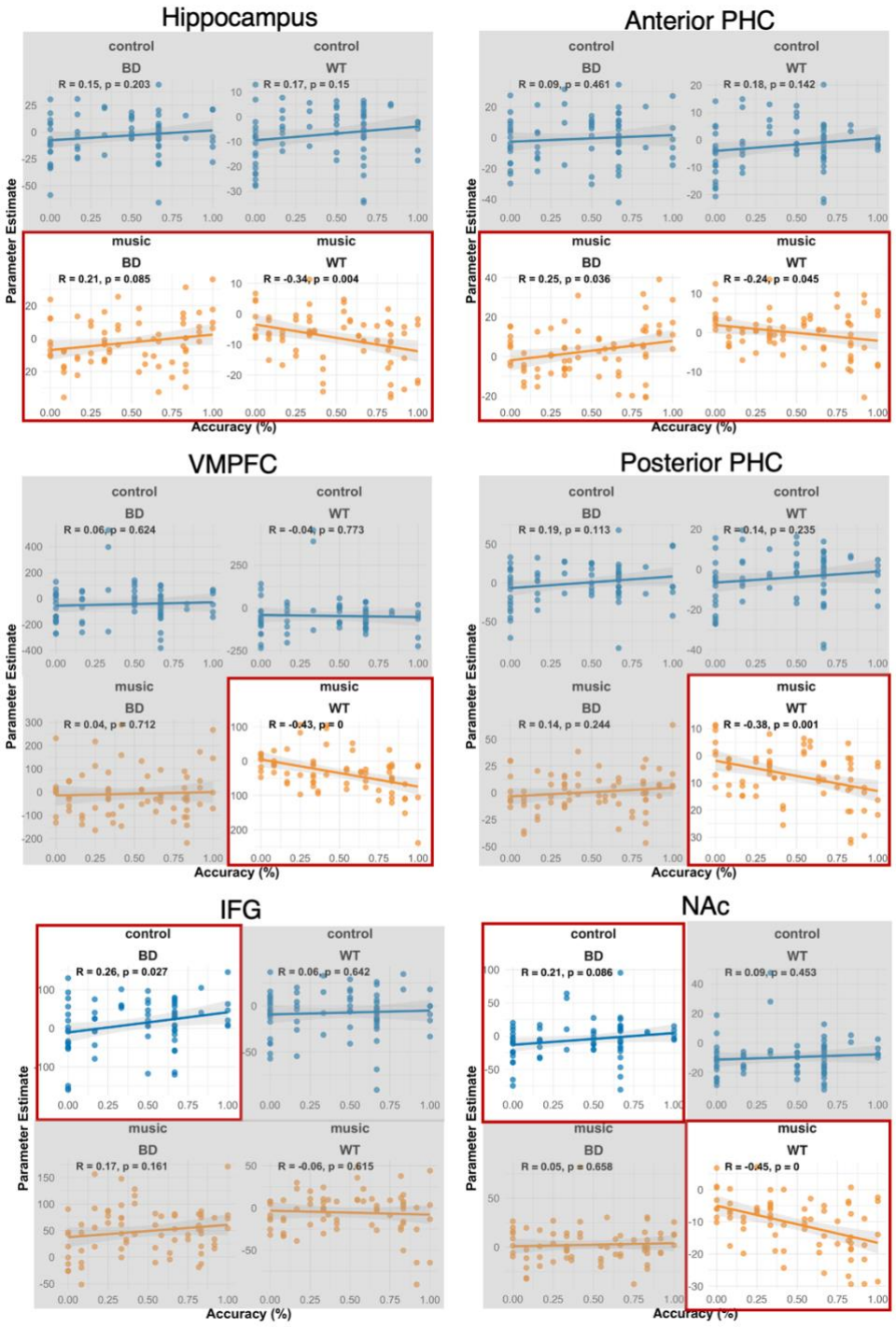

Figure 8 Correlation Between ROI Activity and Behavioral Performance

This figure shows correlations between activity in ROIs (MTL, VMPFC, striatum) and retrieval performance. Positive correlations were found for boundary positions during the music condition

in MTL. Negative correlation was found between BOLD activity during within-sequence items viewing and subsequent retrieval performance in MTL, VMPFC and striatum, only for music condition. IFG showed strong positive boundary activity and retrieval performance for control.

3.4. Finite Impulse Response Analysis Results

Our correlation analyses revealed that MTL and IFG activity at boundaries was correlated to memory performance, with MTL showing particularly strong relationships in the music condition and IFG in the control condition. To further investigate the neural dynamics at sequence boundaries, especially if responses to these events shift earlier or later in time (reflecting more predictive or prolonged processing of the event boundary), I conducted a Finite Impulse Response (FIR) GLM analysis using an 8-second window following the onset of the last picture in each sequence. This analysis aimed to capture the time course of BOLD responses within regions of interest. While vmPFC did not show boundary-specific correlations with behavior, I included it as a comparison ROI given its established role in schema-based learning (van Kesteren et al., 2010) and its significant within-sequence effects in our previous analyses (Ezzyat & Davachi, 2021). This FIR analysis served three key purposes: 1) to characterize potential shifts in boundary responses that might reflect predictive or prolonged processing, 2) to examine how these temporal dynamics might differ between music and control conditions, and 3) to track how boundary responses evolve across learning (early, middle, and late encoding runs). By comparing the time courses across these regions of interest, we can better understand how music might modulate not just the magnitude, but also the timing of boundary-related neural processes.

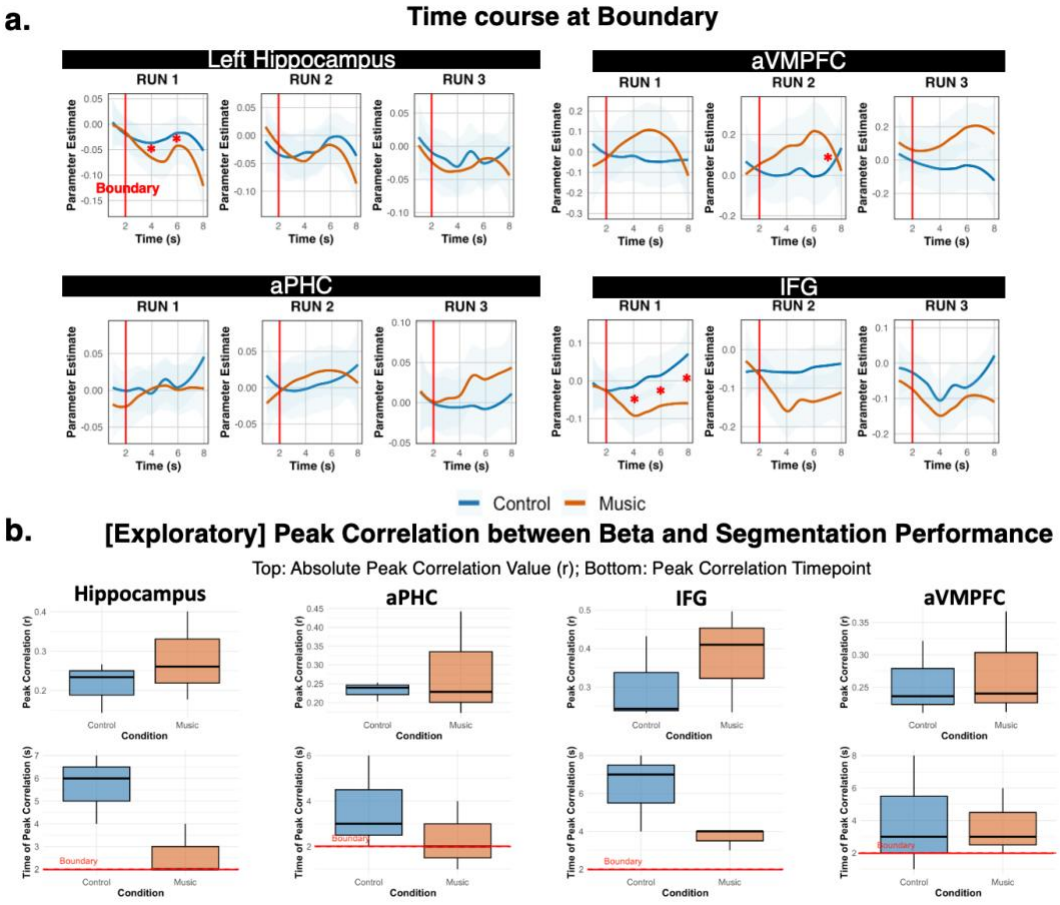

Figure 9 Post-Boundary Activity and Association with Segmentation Performance

(4) *Time Courses of BOLD Responses*: This shows the time courses of BOLD responses in the MTL, IFG and VMPFC following the onset of the last picture in each sequence. Music condition led to a stronger reduction in BOLD activity, particularly during early encoding runs, in MTL and IFG. **(b)** *Correlation Between Post-Boundary BOLD Responses and Segmentation Performance:* This subfigure displays the peak correlations between post-boundary BOLD responses at each 1s time bin and segmentation performance in the music and control conditions. Higher correlations between BOLD activity and segmentation performance were observed in the music condition, with peak correlations occurring closer to the boundary onset, especially in MTL and IFG.

The time courses of BOLD responses are shown in **Figure 9**.a. I used t-tests to compare control and music conditions at each time point. In the hippocampus and inferior frontal gyrus (IFG), I observed a divergence in neural activity between music and control conditions, particularly during the early encoding run (Run 1). The music condition elicited a stronger reduction in BOLD activity shortly after boundary onset in both regions. In the hippocampus, this reduction was evident during Run 1 at Timepoint 4 ($t(32) = 2.362$, $p = .024$, $pFDR = .092$). This pattern suggests that music might facilitate the initial detection of event boundaries (reflected in reduced FIR response) while the hippocampus' broader role in binding and organizing temporal information remains critical for successful memory formation when music is accompanied (reflected in the positive correlation with performance). A similar pattern was observed in the IFG, with an even more pronounced reduction at the same timepoint ($t(32) = 3.251$, $p = .002$, $pFDR = .011$).

The anterior ventromedial prefrontal cortex (vmPFC) demonstrated a contrasting pattern, showing significant increased BOLD activity in the music condition compared to control, particularly during the middle encoding run (Run 2). This increased activation peaked 7 seconds after the onset of the last picture ($t(32) = -3.245$, $p = .003$, $pFDR = .020$).

To further explore the functional engagement of the observed neural patterns, I conducted an exploratory analysis correlating the BOLD responses at each time point with participant'' segmentation performance within the same run. While this analysis is exploratory, it can reveal whether neural responses to boundaries become more predictive over time (shifting earlier relative to the boundary) in the music condition compared to control, potentially indicating learning of the temporal structure provided by musical context. Additionally, examining these

time-resolved correlations can help identify when neural activity is most strongly linked to successful boundary detection, and whether this temporal relationship differs between music and control conditions.

As shown in **Figure 9**.b, the peak correlation values revealed that the hippocampus and IFG exhibited higher correlations between BOLD activity and segmentation performance in the music condition compared to the control condition. Notably, in the music condition, these peak correlations tended to occur closer to the boundary onset (**Figure 9**.b: bottom plots), suggesting more efficient and/or earlier boundary processing when musical context was present. This temporal shift may indicate that boundary-related neural processing begins at or slightly before the actual boundary when supported by musical structure.

In the anterior vmPFC, the correlation pattern differed, with less difference observed between the control and music conditions. However, it is important to note that these findings are exploratory, and few correlations reached statistical significance.

Overall, the FIR analysis revealed distinct temporal patterns of neural activity in key regions involved in boundary detection and segmentation. The hippocampus and IFG showed a stronger BOLD signal reduction in the music condition during early encoding, while the anterior vmPFC exhibited increased activation in the music condition during middle encoding. The exploratory correlation analysis suggests that music may facilitate more efficient segmentation processing, particularly in the hippocampus and IFG, as indicated by the timing of peak neural-behavioral correlations. These findings provide insights into the temporal dynamics of neural processes involved in sequence segmentation and the potential modulatory effects of musical context.

3.5. Functional Connectivity Results

3.5.1. Seed-based gPPI and ROI-to-ROI Connectivity

A generalized psychophysiological interaction (gPPI) analysis using the CONN toolbox was conducted to examine how musical context influences network-level interactions during sequence learning. Seed-based connectivity results are reported in ***Table 6*** and ***Table 7***, containing significant clusters found to be connected with each seed region differently for: 1) music vs. control overall; 2) music vs. control at event boundary; and 3) music vs. control within the sequence.

Table 6 Seed-Based Functional Connectivity Results for Music > Control

Music > Control							
Cluster (x,y,z)	*size*	*size p-FWE*	*size p-FDR*	*size p-unc*	*peak p-FWE*	*peak p-unc*	*Regions*
ENCODING OVERALL							
				Amygdala			
32 -74 +50	195	0.00000	0.00000	0.00000	0.24061	0.00000	LOC
54 +30 +22	48	0.05188	0.01415	0.00659	0.99999	0.00008	MFG, IFG
52 +10 +36	41	0.10435	0.02196	0.00142	1.00000	0.00010	Precentrual Gyrus
				Hippocampus			
-08 -54 54	70	0.00684	0.00524	0.00009	0.91482	0.00001	Precuneus
				aPHC			
52 -50 +36	83	0.00247	0.00194	0.00003	0.82810	0.00001	IFG
2 +30 +36	79	0.00346	0.00194	0.00005	0.98746	0.00002	Retrosplenial Corte
				pPHC			
4 -88 4	117	0.00018	0.00013	0.00000	0.99694	0.00003	Lingual Gyrus, Occ Pole
				Caudate			
60 -28 12	46	0.06912	0.04191	0.00094	0.99943	0.00005	Planum Temporale
-56 -28 +10	42	0.10230	0.04191	0.00142	0.89618	0.00001	Planum Temporale
AT BOUNDARY							
				Hippocampus			
30 22 56	88	0.00117	0.00060	0.00002	0.99481	0.00003	SFG/SMA
22 -88 -4	65	0.00983	0.00591	0.00013	0.99980	0.00005	Fusiform Gyrus
-8 -88 -30	41	0.08877	0.03647	0.00114	0.99176	0.00002	Cerebellum
				aPHC			
-4 -62 32	103	0.04961	0.04477	0.00022	0.87958	0.00001	Precuneus

Table 7 Seed-Based Functional Connectivity Results for Control > Music

Control > Music							
Cluster (x,y,z)	*size*	*size p-FWE*	*size p-FDR*	*size p-unc*	*peak p-FWE*	*peak p-unc*	*Regions*
ENCODING OVERALL							
				NAc			
16 +40 -10	50	0.04676	0.02012	0.00063	0.99995	0.00007	Orbitofrontal Cor
-14 +40 +0	44	0.08372	0.02012	0.00115	0.94841	0.00001	DLPFC
				aPHC			
52 -50 +36	83	0.00247	0.00194	0.00003	0.82810	0.00001	AG
2 +30 +36	79	0.00346	0.00194	0.00005	0.98746	0.00002	Paracingulate Gyr SFG
24 +50 +26	43	0.09241	0.02710	0.00128	0.99996	0.00005	DLPFC
-28 -64 -30	34	0.22486	0.04746	0.00335	0.99206	0.00003	Cerebellum
AT BOUNDARY							
				aVMPFC			
-50 -40 30	42	0.08645	0.03627	0.00113	0.99942	0.00004	SMG
				pPHC			
50 42 0	44	0.06486	0.04934	0.00082	1.00000	0.00014	IFG
WITHIN SEQUENCE							
				Hippocampus			
-58 -30 4	52	0.03174	0.01981	0.00041	1.00000	0.00008	STG
6 50 -4	47	0.06068	0.03897	0.00082	1.00000	0.00017	Anterior Cingulate Gyrus

The control condition demonstrated stronger connections between the medial temporal lobe (MTL) and striatum with prefrontal cortex regions such as the dorsolateral prefrontal cortex (DLPFC), cingulate gyrus, and inferior frontal gyrus (IFG). In contrast, in the music condition, MTL regions including the hippocampus, amygdala, and parahippocampus tended to show stronger connectivity with the visual information processing stream, including lateral occipital cortex, lingual gyrus, fusiform gyrus, as well as the precuneus, especially the retrosplenial cortex. This pattern suggests that musical context may enhance the integration of visual information into memory circuits.

ROI-to-ROI gPPI connectivity analysis revealed two primary clusters showing significantly different connections between music and control conditions during sequence learning: one cluster (shown in red in **Figure 10**) exhibited significantly higher connectivity during the music condition ($F(2,33) = 8.45$, $p = .001$, $pFDR = .021$). These clusters mainly connected the MTL and the ventromedial prefrontal cortex (vmPFC). For example, the posterior vmPFC and right hippocampus showed increased functional connectivity for the music condition during sequence encoding ($t(34) = 3.3$, $p = .002$, $pFDR = .058$). Another cluster (shown in blue in **Figure 10**) demonstrated increased connectivity in the control condition ($F(2,33) = 6.75$, $p = .003$, $pFDR = .035$), mainly located in the left hemisphere connecting frontal regions such as the IFG with the inferior parietal lobule, a region highly associated with syntax and regularity learning.

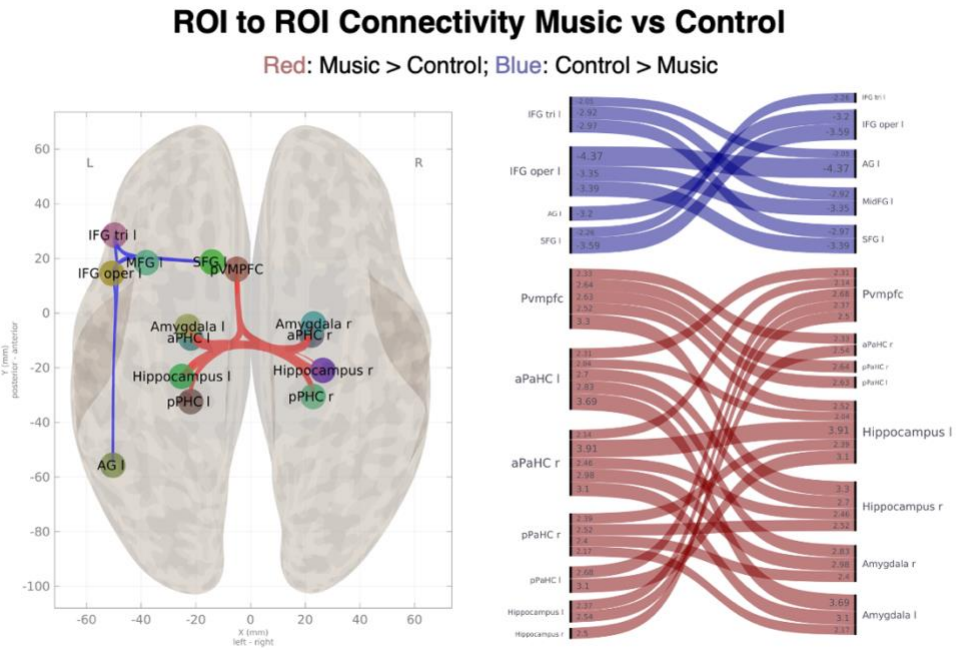

Figure 10 ROI-ROI Connectivity during Encoding

This figure shows functional connectivity among ROIs for the music and control conditions. Music enhanced connectivity between the MTL and vmPFC. Control showed stronger IFG-parietal connectivity, linked to rule learning and syntactic processing.

3.5.2. Connectivity and Behaviors

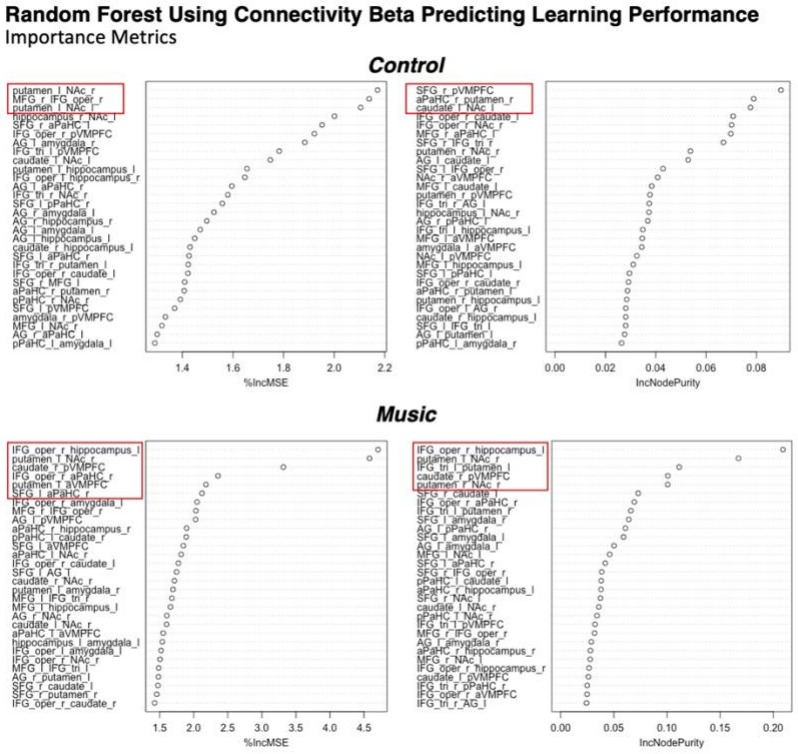

Figure 11 Random Forest Importance Metrics

Random forest model was used to illustrate the relationship between ROI-to-ROI connectivity and sequence learning accuracy. This figure shows the importance metrics of the music and control random forest model.

As with my prior analyses, I tested for associations between the connectivity profiles and with learning behaviors. Due to the relatively small sample size but large number of multiple

comparisons (region-region interactions), I employed a random forest model to predict sequence learning outcomes (retrieval accuracy) using ROI-to-ROI connectivity beta values as features.

The control condition model resulted in a negative percentage of variance explained (-27.52%), suggesting that the model's predictions are worse than simply using the mean accuracy as a constant predictor for all observations. In contrast, the music condition model resulted in squared residuals equal to 0.09, and the model explained 9.3% of the variance in accuracy scores. These results suggest that the random forest model was able to capture some of the relationship between the connectivity beta values and memory accuracy only in the music condition. Of interest is which connectivity relationships underly the model performance when attempting to fit behavior. **Figure 11** shows the importance metrics for both models and reveals that the music model found more predictive features, such as the IFG-MTL and IFG/vmPFC-striatum connectivity.

These findings collectively suggest that music modulates functional connectivity patterns during sequence learning, particularly enhancing connections between the MTL and visual processing regions, as well as between the MTL and vmPFC as well as between the PFC and the striatum. The music condition also appears to facilitate a more predictive relationship between brain connectivity and memory performance, highlighting the potential role of music in shaping neural networks involved in sequence learning and memory formation. These findings expand our understanding of how musical context shapes not just regional activity but also the broader neural networks supporting temporal sequence learning.

3.6. Representational Similarity Analysis

3.6.1. Pattern Similarity for Hippocampus Subfields for Different Item-Item Relationship

The primary aim of this analysis was to understand how hippocampal neural patterns, along with those of its subregions, contribute to encoding the positional and temporal relationships between items with music context. After conducting General Linear Models (GLMs) that treated each item as a stimulus of interest, I extracted the beta patterns for each item from both the left and right hippocampal body masks. These betas served as representations of hippocampal neural patterns during learning. For each run, I computed the neural pattern similarity for all possible item pairs using Pearson's correlation coefficient. **Figure 2**.c presents the averaged representational similarity matrix for 36 items in the right hippocampal body (#1-#24 were learned with music). Both music and control conditions exhibited a similar pattern: hippocampal patterns for items within the same sequence were positively correlated, whereas patterns for items from different sequences but learned in the same block were negatively correlated.

While no qualitative differences between the music and control conditions was observed in the hippocampal body, the next step was to investigate how hippocampal subfields contribute to learning sequential item relationships. Using manually traced subfield masks for each subject (**Figure 2**.b), I compared the music and control conditions in terms of neural pattern similarity across three types of item relationships: 1) items within the same sequence (*within sequence*), 2) items across sequences but learned within the same block (*across boundary*), and 3) items from different blocks (*across block*). **Figure 12** illustrates the pattern similarity for these three types of item relationships across ten hippocampal subregions. A repeated-measures ANOVA was conducted to determine whether the region's neural similarity was modulated by 1) context

condition, 2) the type of sequential relationship, and 3) their interaction. Tukey's HSD post-hoc pairwise comparisons were then performed to further explore differences between the music and control conditions for each type of sequential relationship.

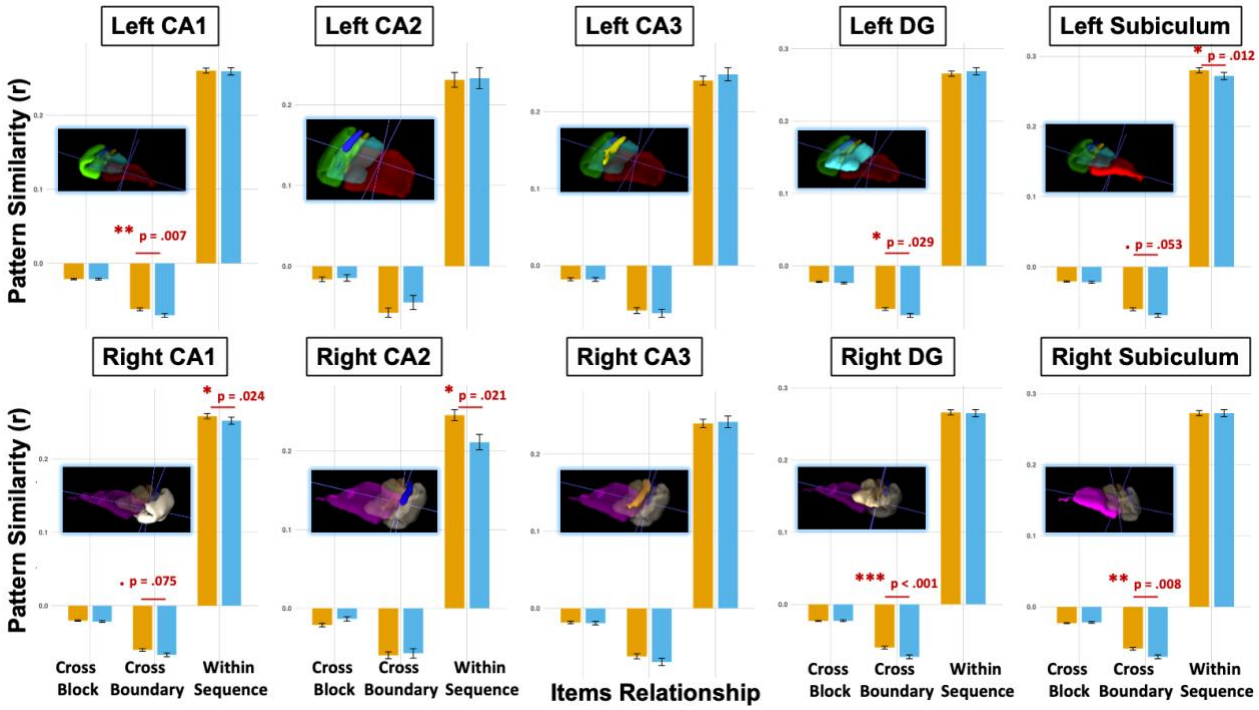

Figure 12 Hippocampal Subfields – Pattern Similarity for Item Relationships

Pattern similarity between items within sequence, across sequences and across block is shown across ten hippocampal subregions. The music condition showed higher within-sequence pattern similarity in CA1, CA2, DG and subiculum, whereas it showed differential cross-sequence pattern similarity in CA1, DG and subiculum.

For the left CA1, the control condition showed a more negative correlation for cross-sequence/within-community items ($t(18,758) = 2.7$, $p = .007$), with a trending interaction effect between context and item relationship ($F(2,18,758) = 2.458$, $p = .086$). In the right CA1, neural pattern similarity was modulated by context $F(1,18,758) = 6.17$, $p = .013$), and there was a

significant interaction between context and sequential relationship type ($F(2,18,758) = 6.151$, $p = .002$). Within the music condition, pattern similarity for items within the same sequence was higher than in the control condition ($t(18,758) = 2.258$ $p = .024$) whereas the cross-boundary items showed trending stronger negative correlation for control condition ($t(18,758) = 1.777$, $p = .075$).

In the right CA2, the ANOVA revealed a trending interaction effect between context and item relationship type ($F(2,18,111) = 2.345$, $p = .096$). Here again, the music condition exhibited higher pattern similarity for within-sequence items than the control condition ($t(18,111) = 2.308$, $p = .021$).

No significant effect was observed for CA3.

The dentate gyrus showed bilateral effects of musical context. Right DG exhibited significant context modulation ($F(1,18,758) = 5.371$, $p = .021$) and trending interaction effects ($F(2,18,758) = 2.809$, $p = .060$). Left DG showed similar trending interactions ($F(2,18,758) = 2.963$, $p = .0517$). Both hemispheres demonstrated less negative correlation for cross-boundary items in the music condition (left: $t(18,758) = 2.184$, $p = .029$; right: $t(18,758) = 3.231$, $p = .001$).

The left subiculum exhibited a strong context effect ($F(1,18,758) = 5.805$, $p = .016$). In the music condition, cross boundary items were less negatively correlated ($t(18,758) = 2.023$, $p = .043$), while within-sequence items showed a trend towards higher positive correlation in this region ($t(18,758) = 1.897$, $p = .058$). A similar pattern was observed in the right subiculum, where context significantly modulated pattern similarity ($F(1,18,575) = 4.657$, $p = .031$). Once again, the music condition demonstrated less negative correlation for cross boundary items compared to the control condition ($t(18,758) = 2.657$, $p = .008$).

3.6.2. Adjacent Items Pattern Similarity and Their Retrieval Accuracy

In this task, participants engaged in a brief retrieval practice after each block (**Figure 13**.b). They were presented with a cue image and instructed to recall the next item in the sequence. The hippocampal pattern similarity between each item and its subsequent item may differ by whether participants had learned the sequential relationship between them. Additionally, I aimed to explore how the music context influenced this relationship.

To investigate this, I utilized the representational similarity matrices generated from previous analyses. I identified the specific items participants were tested on in each block/run and extracted the pattern similarity between each item and its subsequent item, excluding trials where the last item in the sequence was tested. This analysis was conducted separately for the left and right hippocampal body.

I applied a mixed-effects linear model to examine how memory accuracy (whether participants remembered or forgot the next item), encoding phase/run, and context condition jointly influenced pattern similarity, treating participants as random effects. Appendix ***Table 12*** and **Table 13** presents the ANOVA results for the two models (left and right hippocampus body).

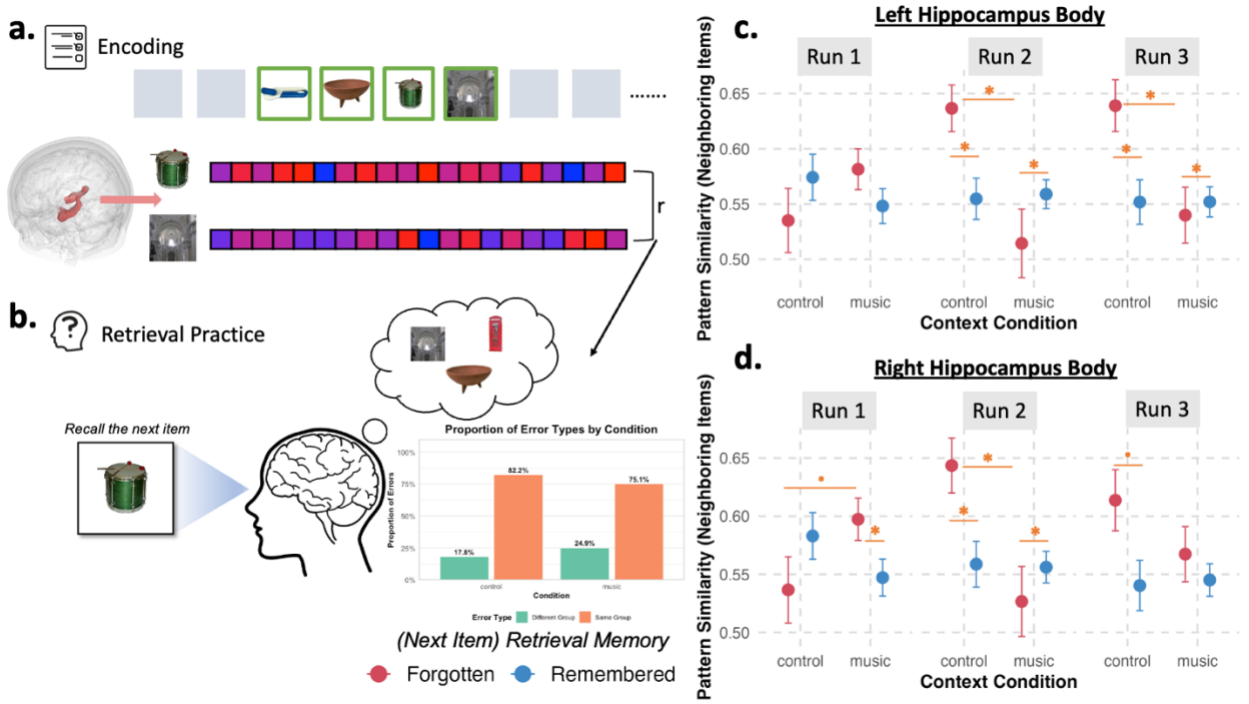

Figure 13 Adjacent Items Pattern Similarity and Retrieval Practice Accuracy

(a)(b) Hippocampus pattern similarity for adjacent items that were tested during retrieval practice was compared for correctly remembered items versus forgotten items. **(c)** shows how pattern similarity between consecutive items in the left hippocampus is influenced by context (music vs. control), encoding run (early, middle, late), and memory accuracy (remembered vs. forgotten). **(d)** illustrates the pattern similarity across encoding runs in the right hippocampus. The similarity pattern was distinct between music and control condition, especially for incorrect items.

For the left hippocampus model, the results indicated that pattern similarity between consecutive items was significantly modulated by the interaction of context and run ($F(2,2404.7) = 5.576$, $p = .004$), as well as by the interaction of context, run, and accuracy ($F(2,2408.3) = 5.686$, $p = .003$). Additionally, there was a trend-level difference in pattern similarity between remembered and forgotten items ($F(1,1272.8) = 2.793$, $p = .094$). Given the large number of

pairwise comparisons, and for exploratory purposes, unadjusted pairwise comparisons were conducted. As shown in **Figure 13**.c, for runs 2 and 3, the music and control conditions exhibited opposing patterns for forgotten and remembered items. In the control condition, hippocampal pattern similarity between neighboring items was higher for forgotten items, whereas in the music condition, the similarity was higher for remembered items (see Appendix **Table 14** and **Table 15** for full statistics).

In the right hippocampus, the mixed-effects model revealed a significant main effect of context ($F(1,2412.1) = 6.251$, $p = .013$), along with a strong *context*run* interaction ($F(2,2404.1) = 6.234$, $p = .002$) and a three way *context*run*accuracy* interaction ($F(2,2408.3) = 5.853$, $p = .003$). **Figure 13**.d illustrates these distinct patterns between conditions, particularly during run 2.

3.6.3. Position-Specific Neural Representations in Hippocampal Body

One insight from above findings is that music provided stronger and more consistent context for encoding the temporal order relationship of items. One hypothesis is that the positional representation inside the hippocampus is more stable for music condition. Binary SVM classifiers were used to distinguish between hippocampal pattern for adjacent positions.

Figure 14 showed the classification accuracy distribution for the hippocampal body (right and left combined), for music and control models. Across all adjacent pairs, classification accuracy was significantly above chance level (50%) for both music and control conditions (music: $t(89) = 11.34$, $p < .001$; control: $t(89) = 5.626$, $p < .001$), indicating distinct neural representations for different sequence positions. Importantly, the music condition showed higher classification accuracy ($M = 58.84\%$, $SD = 7.4\%$) compared to the control condition ($M = 54.1\%$, $SD = 7\%$; paired t-test, $t(89) = 4.286$, $p < .001$, $d = .452$). This enhanced position

discrimination was particularly pronounced for early sequence positions (1→2 transition: music M = 64.6%, control M = 51.9%, t(29) = 9.42, p < .001), shown in **Figure 14**, suggesting that music helps establish more distinct position representations early in the sequence.

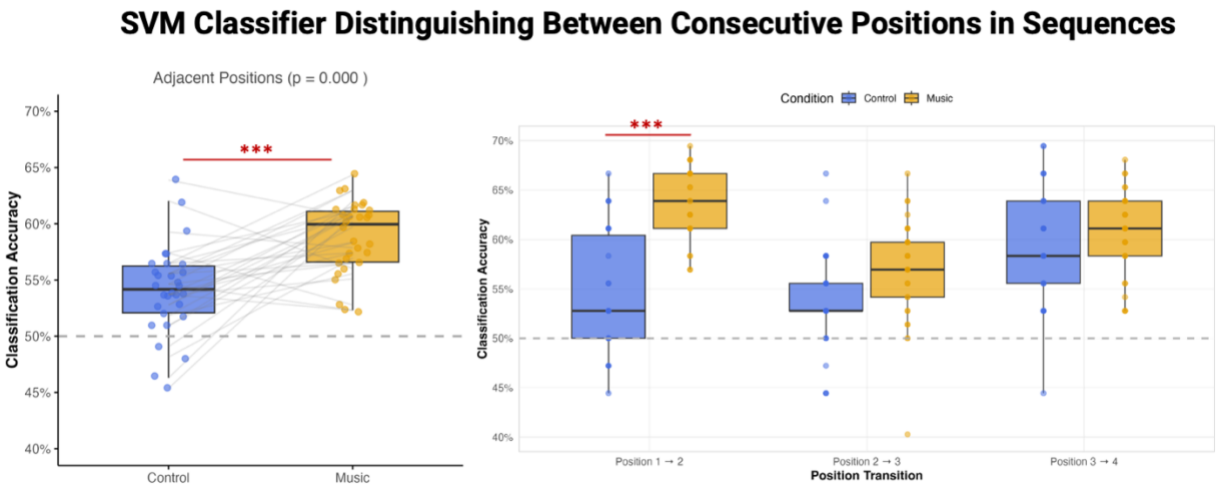

Figure 14 Classification: Using Hippocampal Neural Patterns to Predict Item Position

Classification accuracy of support vector machine (SVM) models trained to distinguish between hippocampal activity patterns from adjacent positions in sequences. *Left*: Overall classification performance for adjacent positions shows higher accuracy in the music condition (orange) compared to control (blue). *Right*: Position-specific analysis reveals that music particularly enhanced the distinctiveness of neural representations for early sequence positions (1→2 transition). Boxplots show the distribution of classification accuracy across subjects, with individual data points overlaid. Dashed line indicates chance level performance (50%). Error bars represent ±1 SEM. *** p < .001.

In summary, our pattern similarity and pattern classification analyses investigated how hippocampal and subfield activity patterns contribute to encoding item positional and temporal relationships. As predicted in the hypotheses, data revealed strengthened items binding and separation in the hippocampus. Music condition had higher within-sequence similarity in CA1, CA2, dentate gyrus (DG) and subiculum, but less negative cross-sequence similarity in CA1, DG and subiculum compared to the control condition. For items tested during retrieval practice, hippocampus body showed opposite patterns on remembered vs forgotten items in later runs for music and control conditions, which might suggest a different mechanism of remembering and forgetting when music is accompanied (see Discussion). Finally, a classification analysis demonstrated the music condition had higher accuracy than control in distinguishing activity patterns of adjacent item positions, particularly for early sequence positions, suggesting music enhances distinct positional representations. These findings together suggest that music enhances the item binding and sequential representation in the hippocampus.

CHAPTER 4. DISCUSSION

4.1. Overview of Key Findings

This study investigated the influence of musical context on learning temporal order relationships of visual items, yielding several important findings. Behavioral results demonstrated higher subsequent retrieval accuracy for visual sequences learned with familiar musical background accompaniment compared to those learned in silence (extending my prior work probing active memorization with music as an aid and with non-musical auditory controls). During encoding, music significantly improved sequence probabilistic boundary detection, suggesting enhanced segmentation of continuous visual streams into events despite random noise (akin to many real-world statistical learning examples). Retrieval tasks inserted during encoding suggested a consistent higher learning rate of sequential relationship of items that were learned in the music condition. Overall, behavioral results indicated music's improving effect on learning temporal order patterns and event segmentation.

Whole-brain fMRI analysis using a General Linear Model (GLM) revealed distinct activation patterns during encoding. As expected, the music condition elicited stronger activation in auditory and motor regions than the silent control, but importantly it also showed evidence of reduced demands on frontal areas which were engaged for sequence processing in the control case. At sequence boundaries, music specifically triggered stronger activation in the amygdala and thalamus compared to the control condition.

Zooming in on the *a priori* ROIs, the correlation between their neural activity and behavioral performance revealed intriguing relationships. MTL activity *at boundaries* positively correlated with later retrieval performance. while its activity *within sequences* negatively

correlated with performance. This might suggest that successful sequence learning was associated with higher MTL sensitivity to boundary points and facilitated processing during within-sequence processing, reflected by the decreased engagement. Similarly, the striatum and vmPFC showed robust negative correlations between within-sequence activity and retrieval performance, indicating that better memory outcomes were associated with reduced activity in these regions during sequence processing within boundaries. These correlation patterns were particularly pronounced in the music condition, suggesting that music might create a more consistent neural state for sequence learning, where the relationship between regional brain activity and behavioral performance becomes more systematic and reliable.

To further elucidate how music affected boundary processing, I employed Finite Impulse Response Analysis, focusing on regions that were boundary-sensitive (MTL, IFG) and vmPFC as a contrast, a region that relates to within-boundary serial memory (DuBrow & Davachi, 2016). I observed a stronger BOLD signal reduction in the hippocampus and IFG immediately after boundary image onset in the music condition. Exploratory neural-behavioral correlations peaked closer to event boundaries in the music condition in the hippocampus and IFG, further corroborating evidence that there is more efficient segmentation processing with music.

Functional connectivity analysis revealed how music modified network interactions during sequence learning. The music condition enhanced connections between the MTL and visual processing regions, as well as between the MTL and vmPFC. In contrast, the control condition showed stronger connections between the IFG and angular gyrus. Random forest models relating brain networks to behavior indicated that the music condition facilitated a more predictive relationship between brain connectivity and memory performance, particularly among the IFG, MTL, striatum, and vmPFC.

Representational similarity analysis of hippocampal subfields, selected based on their established role in sequence encoding and pattern separation, revealed stronger within-sequence similarity for items learned with music, particularly in CA1, dentate gyrus, and subiculum. These regions showed reduced representational repulsion for cross-boundary items in the music condition compared to control. During encoding, higher hippocampal pattern similarity for adjacent items corresponded with successful later retrieval in the music condition, while the control condition showed an inverse pattern. Lastly, when using hippocampal body neural patterns to predict item position in adjacent pairs, classifiers showed higher accuracy in music condition, suggesting more consistent positional representation in the hippocampus with music context. These findings together suggest that music may serve as a shared contextual feature that facilitates memory through distinct encoding mechanisms for item relationships.

Collectively, the results of this study demonstrate that musical context can enhance temporal order memory and statistical learning of visual sequences, potentially through facilitated event segmentation and temporal contextualization. The observed modulation of activity in memory-related regions and alterations in functional connectivity patterns provide a foundation for understanding the neural mechanisms through which music may enhance learning behaviors, which will be examined in detail in the following section.

4.2. Music Enhances Boundary Detection

This study's findings suggest that music can play a crucial role in enhancing mnemonic boundary detection, at least for stimuli processed in parallel within visual processing streams. By pairing distinct music pieces as background stimuli with different visual sequences nested within a continuous presentation stream, the experimental design introduced an additional contextual layer to the learning process. While previous research has established that changes in context

often create event boundaries (DuBrow & Davachi, 2013), this current experimental design tested whether music could provide supplementary boundary cues, even when participants were not explicitly informed about the association between music and item sequences. The observed gradual improvement in boundary detection during encoding indicates that participants did not solely rely on music for this task. Nevertheless, elevated segmentation performance in the music condition suggests musical contextual cues can provide parallel boundary information that converges with ongoing statistical learning processes in areas like the PFC and MTL, aiding the brain in segmenting the concurrent visual streams in memory.

Whole-brain univariate analysis revealed two distinct patterns: during sequence encoding, activity in PFC and visual systems was reduced in the music condition, potentially indicating more efficient processing of sequential information with music providing a structured context. At sequence boundaries in particular, the music condition showed enhanced activation in the inferior frontal gyrus (IFG), thalamus, and amygdala compared to the control. Music processing inherently involves ongoing prediction and anticipation, with musical changes or errors eliciting prediction-violated signals in the brain (Koelsch et al., 2019; Salimpoor et al., 2013). IFG and amygdala have been previously associated with online music prediction, reward-based prediction error processing, and emotion perception during listening (Asaridou et al., 2016; Koelsch, 2014). In the current design, when the music changed, it violated the ongoing prediction of listeners (since they knew the songs so well), particularly during early learning trials when participants had not yet learned the temporal extent of each musical segment. These signals could promote the memory model update and help mark event boundaries through changes in the context trace (Rouhani et al., 2020; Zacks et al., 2011). These context boundaries, in turn, appear to facilitate better learning of sequential item relationships (Pettijohn et al., 2016), consistent with the

observed behavioral results showing improved sequential relationship learning within the same event boundary, paralleling a previous study (Ezzyat & Davachi, 2011). Interestingly, the neural results showed that boundary-related activity in the medial temporal lobe (MTL), specifically the hippocampus and anterior parahippocampal gyrus, positively associated with subsequent sequence retrieval performance, but this was only significant in the music condition. This again suggests that boundaries in musical context cues not only improve correct segmentation but also enhance encoding of temporal relationships.

Finite Impulse Response (FIR) analysis provided additional insights into brain activity patterns responding to boundary items (here the final item of each sequence). Previous studies have indicated significant changes in brain regions such as the hippocampus (DuBrow & Davachi, 2016) and IFG (particularly for music; Sridharan et al., 2007) at event boundaries. If boundaries are more clear and better predicted in memory, one would expect earlier and stronger neural responses. The FIR results revealed evidence that this was the case in the music condition, which amplified and shifted the temporal profile of neural responses to boundaries, especially in the hippocampus and IFG ROIs.

In particular, The hippocampus exhibited an initial activity decrease at boundaries during early encoding in both conditions, with a more pronounced reduction in the music condition. This pattern may reflect more efficient event segmentation when musical structure provides clear temporal cues. The IFG showed differential responses between conditions: while the control condition showed minimal immediate post-boundary changes, the music condition demonstrated a marked activity reduction, possibly reflecting parallel segmentation processing across modalities. The ventromedial prefrontal cortex (vmPFC) showed immediate increased responses post boundary in the music condition, emerging relatively later in the time course. Given the

extended nature of the BOLD response, these temporal differences should be interpreted in relative terms - the vmPFC's later-emerging differential response suggests its role may be more involved in subsequent sequence processing than immediate boundary detection.

Exploratory analysis correlating FIR responses with behavioral event segmentation performance during encoding revealed that in both the hippocampus and IFG, peak correlations between beta values and behaviors were not only higher in the music condition but also occurred closer to boundaries compared to the control condition. This implies that predictable music may provide more predictable boundaries, allowing these regions to react to boundaries earlier than in the control condition. Furthermore, the IFG's reaction to boundaries might not only relate to the background music but also transfer this musical boundary information to visual segmentation, providing support for the cross-modal knowledge transfer hypothesis, which will be explored further in the next section.

In conclusion, these findings demonstrate that musical context significantly enhances boundary detection in visual sequence learning. This enhancement is reflected in improved behavioral performance, increased activation in brain regions that relates to prediction error, and more efficient neural responses to boundaries in the hippocampus and IFG. These results not only expand our understanding of event segmentation but also highlight the potential of cross-modal cues in facilitating learning and memory processes.

4.3. Music Schema Can Transfer to New Learning of Sequential Structure

While the previous section highlighted music's role in boundary detection enhancement, , the benefits of music extend beyond this function. **Figure 8** demonstrates that an association between subsequent sequential recall performance and BOLD activity in several of the a priori ROIs emerged during within-sequence item viewing in the music condition. This indicates that

music's effect also occurs at non-boundary moments, providing mnemonic support for item relationship understanding and encoding.

One theoretical framework inspiring this study, which could help explain this relationship, is schema theory. Schemas have been proposed as existing memory frameworks which can improve new learning through crosstalk between the prefrontal cortex and encoding of new events such as my sequential stimuli by the medial temporal lobe (MTL) (Gilboa & Marlatte, 2017; van Kesteren et al., 2012). Building on my previous study, which demonstrated enhanced functional connectivity between PFC and MTL when visual shape sequences were learned with more-schematic music (Ren & Brown, under review), the current experiment employed highly familiar and schematic compositions. This design allowed examination of how sequential structure in familiar/regular music might engage interactions between the MTL and other regions, including the vmPFC , a region reviewed in many studies to help process new memory formation relative to [non-music] schemas (van Kesteren et al., 2013), and the striatum, which contributes to music temporal processing (Grahn & Rowe, 2013) and, thus, the structural information which makes up a music composition's schema.

Functional connectivity analyses revealed distinct networks emerging for control versus music conditions during encoding. Encoding the sequences in silence engaged frontal-parietal connectivity (**Figure 10**), consistent with prior literature – in particular, both the inferior frontal gyrus (IFG) and the angular gyrus have been shown to be strongly associated with language learning and artificial grammar learning (Karuza et al., 2013; Skosnik et al., 2002). The IFG-angular gyrus connectivity has been particularly linked to high-level language grammar processing rather than simple symbol repetition detection (H. Kim et al., 2022), suggesting its importance in processing hierarchical sequence "rules" or organization. That type of nested

structure was prominent in the current statistical learning task demands, because recurrent sequences exist embedded with lure stimuli as well as noise (occasional randomization in order rendering the relationships probabilistic). It was thus intuitive that this functional connectivity relationship was identified in the control condition, given the demands on extracting hierarchical regularity (not only which items are grouped together but how they are sequenced).

In contrast, the music condition was characterized by dominant MTL-vmPFC connectivity. Given MTL's core role in this type of mnemonic sequencing task, this connection may reflect schema transfer processes from the vmPFC to the MTL, as described in previous models (e.g. van Kesteren et al., 2010). Such connectivity could mediate how memory of the music compositions provides associative scaffold for new learning of the visual sequential rules. Indeed, prior studies have suggested that MTL-vmPFC connections support how prior memory modulates or even biases encoding and decision-making (Gluth et al., 2015; van Kesteren et al., 2010). One study specifically showed that this connectivity was associated with how well prior experience was transferred to new learning (Gerraty et al., 2014). Thus, the shift in network connectivity from IFL-AG to MTL-vmPFC triggered by music aligned with my theoretical framing for this study that music might facilitate the application of existing schemas to new sequential learning, prompting us to further investigate how these network changes relate to behavioral performance.

Random forest modeling revealed differential relationships between network connectivity and behavioral performance across conditions. While connectivity patterns in the control condition did not significantly predict behavior, the music condition showed more robust predictive relationships. This differential predictive power warrants careful interpretation. Rather than suggesting that connectivity is unimportant for learning without music (as substantial prior

literature demonstrates the crucial role of neural connectivity in sequential and associative memory formation; e.g., DuBrow & Davachi, 2016; Karuza et al., 2017)), this finding might reflect several possibilities. The control condition may involve more complex or variable connectivity-behavior relationships across participants, or may engage multiple parallel neural processes not fully captured by pairwise connectivity measures. Additionally, participants might employ more diverse learning strategies without musical structure.

On the other hand, the predictive music model suggests that in the music condition, learning performance relied on more consistent interactions between different regions. Examining the model's "importance" metrics, I found that the most crucial connections in the music model for predicting behavior were the interactions between the prefrontal cortex (PFC) and striatum, and between the PFC and MTL. This observed interactions between the PFC (including IFG and vmPFC), striatum, and MTL in the music condition might suggest a complex mechanism by which musical schemas facilitate prediction-related learning. This process likely involves several interconnected components. Music, with its inherent rhythmic and melodic structure, engages the striatum in temporal prediction (Grahn & Rowe, 2013). This may create a temporal framework onto which visual sequences can be mapped, enhancing the encoding of item order and timing. The PFC-striatum interaction observed in current study aligns with known reward prediction error circuits that motivates memory and decision making (Bialleck et al., 2011; Frank et al., 2019). In the context of current task, "rewards" might be conceptualized as successful predictions of upcoming sequence elements. Music may provide a predictable backdrop against which visual sequence predictions become more salient, potentially enhancing learning through prediction error signals. Lastly, the involvement of the MTL's connection with the vmPFC suggests a pathway through which musical schemas may facilitate the rapid

integration of new sequence information into existing knowledge structures that is consistent with past evidence for the schema theory(van Kesteren et al., 2010) and the complementary learning systems theory (which posits that the MTL and neocortex (including the vmPFC) interact to support both rapid learning and gradual integration of new information (McClelland et al., 2020)). Schema and complementary learning systems likely work in concert to influence memory in conditions like I created in this design, with musical schemas providing a rich, predictable context that engages multiple cognitive processes simultaneously. The result is a learning environment where prediction, attention, and memory systems are optimally engaged, facilitating the encoding and integration of new sequential information.

In future studies, a logical next step to further elucidate the role of mechanisms such as the prediction-guided learning could be to investigate how different levels and types of music predictability (e.g., rhythmic regularity, melodic predictability) or prediction errors from the music would modulate organization of this network. I discuss this further in the Future Directions section under section 4.5.

Taken all together, the functional connectivity analysis provided evidence that music recruits a broader brain network, with these regions collectively providing supportive input to the hippocampus to promote sequential learning. In the next section, I examine the role of the hippocampus more closely, discussing why, with the help of musical context cues, it might behave differently to encode sequential relationships and temporal distances between items.

4.4. Music as Context Helps Bind and Dissociate Events

Having established that musical context aids event segmentation through contextual boundaries and enhances sequential learning by providing a familiar structural schema, this section examined whether these benefits led to more efficient hippocampal processing of

temporal relationships and item order. The analyses revealed four key findings: enhanced pattern similarity for within-sequence items, more efficient pattern separation for cross-boundary items, distinct mechanisms for successful versus unsuccessful memory encoding between music and control conditions, and more consistent positional coding in the hippocampus in the music condition.

Previous research has demonstrated the hippocampus's dual role in both binding events together and dissociating them (Davachi & DuBrow, 2015; DeVito & Eichenbaum, 2011). This duality is particularly crucial for processing continuous experiences, where the hippocampus must simultaneously bind related events while separating distinct episodes (Ranganath & Hsieh, 2016). The current study corroborated and extended these findings: hippocampal pattern similarity showed positive correlations for within-sequence items but negative correlations for cross-boundary items (**Figure 2**). These findings further support the hippocampus's role in binding temporally proximate items while dissociating items that, despite temporal proximity, belong to different sequences.

Analysis of hippocampal subregions revealed enhanced binding and dissociation effects in the music condition. CA1, CA2, and subiculum showed higher pattern similarity for within-sequence items during music presentation. These findings aligned with and extend current understanding of how hippocampal subregions contribute to sequential learning and event segmentation. CA1's role in integrating temporal and contextual information (Eichenbaum, 2014) may be enhanced by music's additional temporal context, as music provides additional temporal context for integration. While historically overlooked in the literature (in part due to past methodological constraints for defining its activity), CA2 has recently been highlighted for its role in supporting temporal and sequential processing by passing signals to CA1 (MacDonald

& Tonegawa, 2021). The observed increased pattern similarity in CA2 for within-sequence items during music condition may reflect enhanced temporal coding, possibly due to music providing a richer temporal context. Lastly, the subiculum, while traditionally viewed as an output structure of the hippocampus, serves a critical role in integrating hippocampal outputs with distributed cortical regions involved in contextual encoding (S. M. Kim et al., 2012). The enhanced pattern similarity for within-sequence items observed in the subiculum during music condition likely reflects its role in strengthening contextual associations at multiple levels: both between sequential items and between these items and their broader temporal context provided by music. This interpretation aligns with recent evidence demonstrating the subiculum's involvement in contextual binding beyond its simple relay functions (Sun et al., 2019). The increased pattern similarity may therefore represent more robust encoding of items within their temporal-musical context, facilitated by the subiculum's extensive connections with both hippocampal processing circuits and cortical regions involved in contextual integration.

For cross-boundary item representational comparisons, decreased repulsion (elimination of dissociation) in CA1, DG, and subiculum was observed during music condition – it likely reflects more efficient pattern separation processes. While this finding might initially seem counterintuitive from a pattern separation perspective, recent theoretical and empirical work suggests this actually reflects more efficient representational organization. One recent study showed that when input patterns share features and therefore have overlapping neural representations, pattern separation can manifest as active “repulsion” between these representations – pushing them further apart in representational space (Wanjia et al., 2021). The reduced negative correlations observed in the music condition may therefore reflect a shift from effortful pattern separation to more efficient representational organization when music provides

clear temporal structure. This pattern also aligns with a prior study showing orthogonalized pattern in DG actually enable better discrimination between similar stimuli (Berron et al., 2016). Moreover, DG also showed its role in detecting context changes (Leutgeb et al., 2007). In particular, while DG showed consistent response to small changes in context, this response is non-linear to the amount of context changes (Schmidt et al., 2012). This might explain why strong musical boundaries might reduce the need for active pattern separation during visual streams processing. Similarly, the finding of decreased repulsion in CA1 might suggest similar mechanism as CA1 actively detects mismatches between current input and stored representations (Duncan et al., 2012). The reduced negative cross-boundary items pattern correlation in music condition might indicate that musical boundaries provide such clear contextual shifts that less active mismatch detection is required.

Interestingly, I did not observe significant changes in representational patterns in CA3 between music and control conditions. While null results should be interpreted with caution, this finding may reflect the distinct computational role of CA3 in the hippocampal circuit. Due to its extensive recurrent collateral connections, CA3 is thought to act as an autoassociative network that generates stable memory representations through pattern completion (Guzowski et al., 2004; Rolls, 2013). This stability may make CA3 representations relatively resistant to contextual modulation by music, allowing it to maintain consistent sequence codes regardless of additional temporal structure.

To understand how hippocampal pattern similarity effects relate to behavior, I examined the relationship between neural patterns and performance on retrieval practice tasks during encoding. The analysis revealed opposing patterns for music and control conditions, particularly for incorrect responses, suggesting different mechanisms for successful and unsuccessful

encoding. While pattern similarity stabilized for correctly remembered items across conditions, incorrect trials showed distinct patterns: the control condition exhibited higher pattern similarity for adjacent items, whereas the music condition showed lower similarity. These distinct patterns suggest there might be different "failure modes" in memory encoding with and without music. In the control condition, forgotten items showed higher pattern similarity with subsequent items, which could reflect errors at retrieval due to overgeneralization in the encoded patterns. This aligns with the idea that hippocampus is critical for distinguishing similar items or items that took place at similar timing or location, to be able to later recognize them and place them in the correct context (e.g., sequence) (i.e., its function in mnemonic pattern separation;(Brown et al., 2014; Ranganath & Hsieh, 2016; Yassa & Stark, 2011)). From this perspective of the control condition, the higher between-item similarity for forgotten trials might reflect a failure in pattern separation and thus higher likelihood of failure to identify the placement between items in memory tests. This aligned with the behavioral trend during retrieval practice (**Figure 13**.b) where people chose more in-sequence lures in the control condition.

Conversely, in the music condition, forgotten items showed weaker representational similarity with subsequent items than remembered items, suggesting a different forgetting mechanism compared to Control. While future investigation is needed for concluding the reason, one speculation is that the forgotten items were ones which failed be strongly bound to the music as a shared context with its next item. This pattern aligns with context-based theories of forgetting, where items become inaccessible due to insufficient contextual binding (Howard & Kahana, 2002; Polyn et al., 2009), and we know from my behavioral data that this binding was beneficial to segmenting the stimuli into their sequential groups and orders (Ren et al., 2024). The increased selection of cross-boundary lures in the music condition suggests that when

musical context failed to effectively bind items, it may have led to a distinct source of forgetting where items became "isolated" (rather than overgeneralized) in memory, such that the neural similarity scores tracked where people failed to bind the within-sequence items to their shared context (a failure which, at fMRI's level of measurement, would driving representational similarity down). This pattern of results aligns with established findings in context-dependent memory research, where degradation of contextual information significantly impairs retrieval processes (Smith & Vela, 2001). Additional evidence for music's role in providing consistent contextual codes and enhancing item binding emerged from the item position classification analysis, which revealed higher model accuracy in the music condition. This suggests that items encoded with musical accompaniment developed more consistent neural representations of their positions and temporal order in the hippocampus, providing further support for music's enhancement of hippocampal encoding processes.

In summary, my RSA findings suggest that music fundamentally alters how hippocampus process sequential information. Music appears to create a more structured encoding environment that changes the balance between pattern separation and completion. The music context led to a stronger item binding for items within the same sequence, as well as more efficient item dissociation and separation when they crossed boundaries. Because of these stronger contextual dependencies music creates, I speculate that when memory encoding was unsuccessful, this was reflected in the music condition by evidence of representations of the items being more contextually isolated (as opposed to evidence in the control condition that memory failures were more strongly associated with weaker stimulus discrimination). While the interpretations are speculative, these striking results add to previous results and indicated representational impacts of the music's effect modulating visual sequence encoding in the MTL.

4.5. Limitations and Future Directions

A primary limitation of this study relates to the visual stimuli selection. The task used over one hundred images from various categories (items, animals, and scenes) to facilitate memorization and discrimination. This diversity was chosen in part to bolster generalizability of the findings to the noisy and multi-modal stimulus encoding scenarios in real life, and in part because it aided participant performance (by making the individual elements of the large visual stimulus set more memorable/discriminable). However, it may also have inadvertently created natural boundaries when image categories changed during stimulus presentation. Although category transitions did not follow any sequential regularity, and current analyses specifically focused on designated sequence boundaries (collapsing any such "endogenous" boundaries), the category changes might have influenced neural responses, particularly in category-selective brain regions (e.g., parahippocampal cortices targeted, among other ROIs, in my analyses). Importantly, this limitation affected both control and music conditions equally, thus is likely not a confound for the main comparisons – nevertheless, future studies could implement rigorous control over stimulus categories and transitions to assess how they might interact with the boundary modulation effects targeted in my results.

The current study focused on comparing music versus control conditions to elucidate music's general effects on visual sequence learning. A crucial next step is investigating how specific musical features contribute to this process. This investigation would be particularly relevant for understanding the network-level results and my hypothesis about prediction-related learning effects. My previous work (Ren et al., 2024) has already developed methodology to examine how musical regularity and familiarity influence sequence learning, finding differential effects between more- and less-schematic music. Building on those findings and the current

results, future studies should systematically manipulate levels of musical predictability and examine the effect of different types and magnitudes of prediction errors. The current dataset, which includes music with varying levels of prediction errors, provides the chance for exploring how these parameters modulate neural dynamics, particularly focusing on striatal connectivity patterns with other brain regions. Understanding how the relationship between musical structure and visual sequence structure affects learning outcomes would further illuminate the mechanisms underlying cross-modal learning enhancement.

This study also opens new avenues for understanding cross-modal sequential learning, demonstrating how structural knowledge in one modality can facilitate learning in another. Future research could investigate other cross-modal contexts, such as motion or language, for memory enhancement. Individual differences in cross-modal learning benefits and the temporal dynamics of cross-modal integration during learning represent important areas for investigation. These studies would help establish whether the benefits observed with music generalize to other forms of structured sensory input.

These findings have significant implications for both clinical and educational settings. In clinical applications, future research should focus on developing music-based interventions for memory disorders and investigating music's role in rehabilitation of sequential learning deficits. The optimization of musical parameters for therapeutic effectiveness requires systematic investigation of how different patient populations respond to various musical features. In educational contexts, future studies should examine how music-enhanced learning protocols can be effectively implemented in academic settings. This includes investigating subject-specific musical enhancement strategies and developing personalized learning approaches that incorporate musical elements.

Future studies should also delve deeper into the neural mechanisms underlying music-enhanced learning. This includes investigating the temporal dynamics of learning enhancement, individual differences in music-induced neural network changes, and the role of attention and arousal in music-enhanced learning. The relationship between musical expertise and cross-modal learning benefits represents another important area for investigation. Understanding these mechanisms will be crucial for optimizing music-based interventions and developing more effective learning strategies.

The integration of behavioral, neural, and applied research approaches will be crucial for realizing the full potential of music-enhanced learning. By addressing these future directions, researchers can build upon the foundation established by this study to develop more effective interventions and applications while advancing our theoretical understanding of how music enhances learning and memory processes.

CHAPTER 5. CONCLUSION

This dissertation presents compelling evidence for the role of music as a potent source of contextual cues for enhancing temporal order memory and learning. By examining the effects of music during a statistical learning task, the study demonstrates that music facilitates both the segmentation of events and the learning of sequential structures, ultimately improving memory performance. These findings align with and extend previous research on how contextual cues modulate memory encoding, but they also introduce new insights into the specific neural mechanisms that support these processes.

The behavioral results clearly indicate that music enhances both sequence learning and event boundary detection. Participants in the music condition exhibited better accuracy in retrieving sequences and were more accurate in detecting event boundaries, which are critical moments where the brain segments continuous information into meaningful units. These findings suggest that music aids in organizing information into more manageable and memorable chunks, allowing for better recall and learning efficiency.

At the neural level, the study reveals distinct patterns of activity in regions associated with memory and learning. The finite impulse response (FIR) analysis highlighted that music accelerates boundary processing, particularly in the medial temporal lobe and inferior frontal gyrus. These regions exhibited faster deflections in BOLD signal following event boundaries in the music condition, indicating that music may facilitate the brain's ability to "close off" one segment and prepare for the next, enhancing overall memory performance. The functional connectivity analysis showed evidence of enhanced communication in an MTL-vmPFC-striatum network when music was present. This supported the hypothesis that music enhances both the

encoding of visual sequences and the integration of this information with existing memory schemas.

One of the striking findings from this study comes from the representational similarity analysis (RSA) of hippocampal subfields. Music was found to modulate the similarity of neural patterns, strengthening the associative binding of items within sequences while enhancing dissimilarity for items across sequence boundaries. Additionally, the hippocampal body exhibited more consistent and predictive neural representations of item positions when sequences were encoded with musical accompaniment. These findings suggest that music facilitates multiple aspects of hippocampal processing: it enhances both binding and dissociation mechanisms, while also promoting more efficient encoding of sequential information.

The findings from this research hold significant implications across multiple domains. From a theoretical perspective, they contribute to our understanding of cross-modal influences on learning and memory, particularly how auditory context (in this case, music) interacts with visual processing and memory systems to improve learning outcomes. This study also advances our knowledge of how the brain dynamically adjusts its activity in response to external cues like music, particularly at critical moments of event segmentation and boundary processing.

Practically, the results suggest that music could be strategically used to enhance learning and memory, especially in settings where understanding sequences or temporal relationships is key, such as in education or skill acquisition. The results also raise the possibility of using music in therapeutic contexts, particularly for individuals with memory impairments such as those with Alzheimer's disease, where enhancing memory for sequences and distinguishing between similar events can be challenging.

In conclusion, this dissertation highlights the important role that music can play in supporting memory and learning, particularly by enhancing event segmentation and improving the encoding of sequential information. The evidence presented here lays the groundwork for further exploration into how cross-modal interactions between music and memory systems can be harnessed to improve cognitive performance in both healthy individuals and clinical populations. These findings have broad implications for educational strategies, therapeutic interventions, and our understanding of the brain's remarkable ability to integrate multisensory experiences to facilitate learning.

APPENDIX. Supplementary Statistical Results

Table 8 ANOVA Results: Segmentation Performance Affected by Context, Presentation Order and Encoding Run

Effect	*Df*	*Sum Sq*	*Mean Sq*	*F value*	*Pr(>F)*
Run	2	3.34	1.6723	28.674	7.74E-13***
Context Condition	1	1.09	1.093	18.742	1.65E-05***
Presentation Order	1	1.11	1.1104	19.038	1.41E-05***
Run:Context	2	0.26	0.1323	2.269	0.104
Run:Order	2	0.02	0.0088	0.15	0.86
Context:Order	1	0.01	0.005	0.086	0.769
Run:Context:Order	2	0	0.0016	0.027	0.974
Residuals	1007	58.73	0.0583		

Table 9 Post-hoc Pairwise Comparisons using Bonferroni's Correction on Segmentation Performance

Run#	*contrast*	*estimate*	*SE*	*df*	*t.ratio*	*p.value*
1	Control Ordered - Music Ordered	-0.1197	0.037	1007	-3.232	0.0076***
1	Control Ordered - Control Shuffled	0.0653	0.037	1007	1.763	0.4693
1	Control Ordered - Music Shuffled	-0.0369	0.037	1007	-0.996	1
1	Music Ordered - Control Shuffled	0.185	0.037	1007	4.995	<.0001***
1	Music Ordered - Music Shuffled	0.0829	0.037	1007	2.237	0.1532
1	Control Shuffled - Music Shuffled	-0.1022	0.037	1007	-2.759	0.0355*
2	Control Ordered - Music Ordered	-0.0451	0.037	1007	-1.217	1
2	Control Ordered - Control Shuffled	0.0691	0.037	1007	1.865	0.3752
2	Control Ordered - Music Shuffled	0.0244	0.037	1007	0.66	1
2	Music Ordered - Control Shuffled	0.1141	0.037	1007	3.081	0.0127*
2	Music Ordered - Music Shuffled	0.0695	0.037	1007	1.877	0.3652
2	Control Shuffled - Music Shuffled	-0.0446	0.037	1007	-1.205	1
3	Control Ordered - Music Ordered	-0.0449	0.037	1007	-1.213	1
3	Control Ordered - Control Shuffled	0.0503	0.037	1007	1.357	1
3	Control Ordered - Music Shuffled	0.014	0.037	1007	0.378	1
3	Music Ordered - Control Shuffled	0.0952	0.037	1007	2.57	0.0619~
3	Music Ordered - Music Shuffled	0.0589	0.037	1007	1.591	0.6721
3	Control Shuffled - Music Shuffled	-0.0363	0.037	1007	-0.979	1

Table 10 ANOVA Results: Retrieval Practice Accuracy Affected by Context and Encoding Run

Effect	*Df*	*Sum Sq*	*Mean Sq*	*F value*	*Pr(>F)*
run	2	3.697	1.849	42.05	< 2e-16***
context	1	0.419	0.4186	9.522	0.00214***
run:context	2	0.014	0.0072	0.024	0.84837
Residuals	517	22.73	0.044		

Table 11 Post-hoc Pairwise Comparisons using Bonferroni's Correction on Retrieval Practice Accuracy

run	*contrast*	*estimate*	*SE*	*df*	*t.ratio*	*p.value*
1	control - music	-0.0459	0.0316	517	-1.453	0.1468
2	control - music	-0.0708	0.0316	517	-2.239	0.0256**
3	control - music	-0.0528	0.032	517	-1.652	0.0992

Table 12 Left Hippocampal Body: Pattern Similarity of Neighboring Items and Retrieval Accuracy

Effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
accuracy	0.2210	0.2210	1	1272.8	2.7933	0.0949.
context	0.1658	0.1658	1	2412.6	2.0952	0.1479
phase	0.0661	0.0330	2	2418.4	0.4175	0.6587
accuracy:context	0.0516	0.0516	1	2413.9	0.6523	0.4194
accuracy:phase	0.1530	0.0765	2	2418.1	0.967	0.3804
context:phase	0.8823	0.4411	2	2404.7	5.5755	0.0038*
accuracy:context:phase	0.8998	0.4499	2	2408.3	5.6864	0.0034*

Table 13 Right Hippocampal Body: Pattern Similarity of Neighboring Items and Retrieval Accuracy

Effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
accuracy	0.0976	0.0976	1	1293.5	1.2836	0.2574
context	0.4750	0.4750	1	2412.1	6.2508	0.0125*
phase	0.1020	0.0510	2	2418.3	0.671	0.5113
accuracy:context	0.2593	0.2593	1	2413.4	3.4122	0.0648.
accuracy:phase	0.1170	0.0585	2	2418.1	0.7696	0.4633
context:phase	0.9475	0.4737	2	2404.1	6.2336	0.0020**
accuracy:context:phase	0.8896	0.4448	2	2407.7	5.8528	0.0029**

Table 14 Post-hoc comparisons: Correct vs Incorrect Retrieved Items Pattern Similarity in Left Hippocampal Body

contrast	estimate	SE	df	t.ratio	p.value
accuracy0 control phase1 - accuracy1 control phase1	-0.0415	0.0341	2415.0022	-1.2149	0.2245
accuracy0 control phase1 - accuracy0 music phase1	-0.0470	0.0328	2407.8838	-1.4340	0.1517
accuracy0 control phase1 - accuracy1 music phase1	-0.0146	0.0298	2414.8266	-0.4900	0.6242
accuracy0 control phase1 - accuracy0 control phase2	-0.1007	0.0425	2403.7035	-2.3714	0.0178
accuracy0 control phase1 - accuracy1 control phase2	-0.0215	0.0327	2415.1369	-0.6573	0.5111
accuracy0 control phase1 - accuracy0 music phase2	0.0216	0.0377	2415.2392	0.5736	0.5663
accuracy0 control phase1 - accuracy1 music phase2	-0.0252	0.0290	2416.4263	-0.8694	0.3847
accuracy0 control phase1 - accuracy0 control phase3	-0.1048	0.0429	2411.3724	-2.4429	0.0146
accuracy0 control phase1 - accuracy1 control phase3	-0.0192	0.0327	2417.1863	-0.5875	0.5569
accuracy0 control phase1 - accuracy0 music phase3	-0.0050	0.0381	2418.4997	-0.1321	0.8949
accuracy0 control phase1 - accuracy1 music phase3	-0.0188	0.0290	2414.3405	-0.6485	0.5167
accuracy1 control phase1 - accuracy0 music phase1	-0.0055	0.0300	2400.2670	-0.1843	0.8538
accuracy1 control phase1 - accuracy1 music phase1	0.0268	0.0265	2401.9896	1.0124	0.3114
accuracy1 control phase1 - accuracy0 control phase2	-0.0593	0.0404	2415.4501	-1.4655	0.1429
accuracy1 control phase1 - accuracy1 control phase2	0.0200	0.0297	2403.1568	0.6732	0.5009
accuracy1 control phase1 - accuracy0 music phase2	0.0631	0.0354	2363.4712	1.7824	0.0748

accuracy1 control phase1 - accuracy1 music phase2	0.0162	0.0256	2401.8800	0.6319	0.5275
accuracy1 control phase1 - accuracy0 control phase3	-0.0634	0.0409	2408.3155	-1.5494	0.1214
accuracy1 control phase1 - accuracy1 control phase3	0.0222	0.0297	2402.4173	0.7474	0.4549
accuracy1 control phase1 - accuracy0 music phase3	0.0364	0.0359	2320.9109	1.0156	0.3099
accuracy1 control phase1 - accuracy1 music phase3	0.0226	0.0256	2402.1408	0.8838	0.3769
accuracy0 music phase1 - accuracy1 music phase1	0.0324	0.0251	2355.9528	1.2909	0.1969
accuracy0 music phase1 - accuracy0 control phase2	-0.0537	0.0393	2410.7644	-1.3679	0.1715
accuracy0 music phase1 - accuracy1 control phase2	0.0255	0.0284	2393.8810	0.8985	0.3690
accuracy0 music phase1 - accuracy0 music phase2	0.0686	0.0340	2415.9840	2.0185	0.0436
accuracy0 music phase1 - accuracy1 music phase2	0.0217	0.0241	2380.5389	0.9012	0.3675
accuracy0 music phase1 - accuracy0 control phase3	-0.0579	0.0397	2412.0533	-1.4559	0.1455
accuracy0 music phase1 - accuracy1 control phase3	0.0278	0.0284	2403.1667	0.9760	0.3292
accuracy0 music phase1 - accuracy0 music phase3	0.0420	0.0345	2418.9123	1.2172	0.2236
accuracy0 music phase1 - accuracy1 music phase3	0.0282	0.0241	2381.7919	1.1697	0.2422
accuracy1 music phase1 - accuracy0 control phase2	-0.0861	0.0369	2412.4949	-2.3335	0.0197
accuracy1 music phase1 - accuracy1 control phase2	-0.0069	0.0246	2404.2104	-0.2786	0.7806
accuracy1 music phase1 - accuracy0 music phase2	0.0362	0.0313	2314.5647	1.1579	0.2470
accuracy1 music phase1 - accuracy1 music phase2	-0.0106	0.0196	2400.1852	-0.5421	0.5878

accuracy1 music phase1 - accuracy0 control phase3	-0.0902	0.0374	2395.3332	-2.4106	0.0160
accuracy1 music phase1 - accuracy1 control phase3	-0.0046	0.0247	2402.4313	-0.1866	0.8520
accuracy1 music phase1 - accuracy0 music phase3	0.0096	0.0318	2245.6078	0.3010	0.7634
accuracy1 music phase1 - accuracy1 music phase3	-0.0042	0.0196	2403.7381	-0.2149	0.8299
accuracy0 control phase2 - accuracy1 control phase2	0.0793	0.0393	2409.1424	2.0182	0.0437
accuracy0 control phase2 - accuracy0 music phase2	0.1223	0.0434	2408.1925	2.8181	0.0049
accuracy0 control phase2 - accuracy1 music phase2	0.0755	0.0363	2414.2769	2.0808	0.0376
accuracy0 control phase2 - accuracy0 control phase3	-0.0041	0.0481	2408.3824	-0.0855	0.9318
accuracy0 control phase2 - accuracy1 control phase3	0.0815	0.0393	2414.5827	2.0741	0.0382
accuracy0 control phase2 - accuracy0 music phase3	0.0957	0.0438	2418.0320	2.1838	0.0291
accuracy0 control phase2 - accuracy1 music phase3	0.0819	0.0362	2415.2038	2.2600	0.0239
accuracy1 control phase2 - accuracy0 music phase2	0.0431	0.0340	2350.9290	1.2672	0.2052
accuracy1 control phase2 - accuracy1 music phase2	-0.0038	0.0237	2404.3924	-0.1589	0.8738
accuracy1 control phase2 - accuracy0 control phase3	-0.0834	0.0397	2406.2423	-2.0987	0.0360
accuracy1 control phase2 - accuracy1 control phase3	0.0023	0.0281	2404.0691	0.0801	0.9361
accuracy1 control phase2 - accuracy0 music phase3	0.0164	0.0345	2305.7639	0.4767	0.6336
accuracy1 control phase2 - accuracy1 music phase3	0.0027	0.0237	2405.5912	0.1123	0.9106
accuracy0 music phase2 - accuracy1 music phase2	-0.0468	0.0305	2314.9345	-1.5343	0.1251

accuracy0 music phase2 - accuracy0 control phase3	-0.1264	0.0438	2413.4859	-2.8843	0.0040
accuracy0 music phase2 - accuracy1 control phase3	-0.0408	0.0340	2358.0646	-1.1992	0.2306
accuracy0 music phase2 - accuracy0 music phase3	-0.0266	0.0391	2417.8753	-0.6811	0.4959
accuracy0 music phase2 - accuracy1 music phase3	-0.0404	0.0305	2331.3158	-1.3259	0.1850
accuracy1 music phase2 - accuracy0 control phase3	-0.0796	0.0368	2403.1747	-2.1631	0.0306
accuracy1 music phase2 - accuracy1 control phase3	0.0060	0.0238	2400.9293	0.2530	0.8003
accuracy1 music phase2 - accuracy0 music phase3	0.0202	0.0311	2261.0959	0.6504	0.5155
accuracy1 music phase2 - accuracy1 music phase3	0.0064	0.0184	2401.2260	0.3497	0.7266
accuracy0 control phase3 - accuracy1 control phase3	0.0856	0.0398	2397.9873	2.1511	0.0316
accuracy0 control phase3 - accuracy0 music phase3	0.0998	0.0441	2404.9065	2.2610	0.0238
accuracy0 control phase3 - accuracy1 music phase3	0.0860	0.0368	2399.5628	2.3388	0.0194
accuracy1 control phase3 - accuracy0 music phase3	0.0142	0.0346	2279.6805	0.4104	0.6815
accuracy1 control phase3 - accuracy1 music phase3	0.0004	0.0238	2398.9990	0.0172	0.9863
accuracy0 music phase3 - accuracy1 music phase3	-0.0138	0.0311	2231.0484	-0.4437	0.6573

Table 15 Post-hoc comparisons: Correct vs Incorrect Retrieved Items Pattern Similarity in Right Hippocampal Body

contrast	estimate	SE	df	t.ratio	p.value
accuracy0 control phase1 - accuracy1 control phase1	-0.0494	0.0348	2414.2732	-1.4197	0.1558
accuracy0 control phase1 - accuracy0 music phase1	-0.0616	0.0334	2408.2739	-1.8423	0.0655
accuracy0 control phase1 - accuracy1 music phase1	-0.0122	0.0304	2414.1329	-0.4017	0.6879
accuracy0 control phase1 - accuracy0 control phase2	-0.1071	0.0433	2404.0154	-2.4715	0.0135
accuracy0 control phase1 - accuracy1 control phase2	-0.0241	0.0333	2414.4416	-0.7238	0.4693
accuracy0 control phase1 - accuracy0 music phase2	0.0106	0.0384	2415.5717	0.2750	0.7833
accuracy0 control phase1 - accuracy1 music phase2	-0.0213	0.0296	2415.9039	-0.7173	0.4732
accuracy0 control phase1 - accuracy0 control phase3	-0.0782	0.0438	2411.7414	-1.7856	0.0743
accuracy0 control phase1 - accuracy1 control phase3	-0.0064	0.0334	2416.7598	-0.1926	0.8473
accuracy0 control phase1 - accuracy0 music phase3	-0.0308	0.0389	2418.6437	-0.7931	0.4278
accuracy0 control phase1 - accuracy1 music phase3	-0.0107	0.0296	2413.5730	-0.3615	0.7178
accuracy1 control phase1 - accuracy0 music phase1	-0.0122	0.0306	2398.3339	-0.3975	0.6911
accuracy1 control phase1 - accuracy1 music phase1	0.0372	0.0271	2402.2871	1.3750	0.1693
accuracy1 control phase1 - accuracy0 control phase2	-0.0577	0.0413	2414.7894	-1.3983	0.1621
accuracy1 control phase1 - accuracy1 control phase2	0.0253	0.0303	2403.4698	0.8353	0.4036
accuracy1 control phase1 - accuracy0 music phase2	0.0600	0.0361	2359.3079	1.6624	0.0966

accuracy1 control phase1 - accuracy1 music phase2	0.0282	0.0262	2402.1730	1.0763	0.2819
accuracy1 control phase1 - accuracy0 control phase3	-0.0288	0.0417	2407.0533	-0.6894	0.4906
accuracy1 control phase1 - accuracy1 control phase3	0.0430	0.0303	2402.7187	1.4166	0.1567
accuracy1 control phase1 - accuracy0 music phase3	0.0186	0.0366	2314.5765	0.5086	0.6111
accuracy1 control phase1 - accuracy1 music phase3	0.0387	0.0261	2402.4325	1.4815	0.1386
accuracy0 music phase1 - accuracy1 music phase1	0.0494	0.0256	2351.2199	1.9300	0.0537
accuracy0 music phase1 - accuracy0 control phase2	-0.0455	0.0401	2411.1548	-1.1356	0.2562
accuracy0 music phase1 - accuracy1 control phase2	0.0375	0.0290	2391.5061	1.2937	0.1959
accuracy0 music phase1 - accuracy0 music phase2	0.0722	0.0347	2416.3048	2.0815	0.0375
accuracy0 music phase1 - accuracy1 music phase2	0.0403	0.0246	2377.3906	1.6395	0.1012
accuracy0 music phase1 - accuracy0 control phase3	-0.0166	0.0405	2412.4351	-0.4093	0.6823
accuracy0 music phase1 - accuracy1 control phase3	0.0552	0.0290	2401.5514	1.9012	0.0574
accuracy0 music phase1 - accuracy0 music phase3	0.0308	0.0352	2418.8060	0.8753	0.3815
accuracy0 music phase1 - accuracy1 music phase3	0.0509	0.0246	2378.7437	2.0718	0.0384
accuracy1 music phase1 - accuracy0 control phase2	-0.0949	0.0376	2411.5486	-2.5205	0.0118
accuracy1 music phase1 - accuracy1 control phase2	-0.0119	0.0252	2404.5362	-0.4735	0.6359
accuracy1 music phase1 - accuracy0 music phase2	0.0228	0.0319	2307.8454	0.7144	0.4751
accuracy1 music phase1 - accuracy1 music phase2	-0.0090	0.0200	2400.4424	-0.4512	0.6519

accuracy1 music phase1 - accuracy0 control phase3	-0.0660	0.0382	2393.1350	-1.7278	0.0842
accuracy1 music phase1 - accuracy1 control phase3	0.0058	0.0252	2402.7374	0.2294	0.8185
accuracy1 music phase1 - accuracy0 music phase3	-0.0186	0.0325	2235.8706	-0.5729	0.5668
accuracy1 music phase1 - accuracy1 music phase3	0.0015	0.0200	2404.0661	0.0761	0.9394
accuracy0 control phase2 - accuracy1 control phase2	0.0830	0.0401	2407.8472	2.0716	0.0384
accuracy0 control phase2 - accuracy0 music phase2	0.1177	0.0443	2408.5541	2.6570	0.0079
accuracy0 control phase2 - accuracy1 music phase2	0.0859	0.0370	2413.5022	2.3202	0.0204
accuracy0 control phase2 - accuracy0 control phase3	0.0289	0.0490	2408.7438	0.5899	0.5553
accuracy0 control phase2 - accuracy1 control phase3	0.1007	0.0401	2413.8406	2.5117	0.0121
accuracy0 control phase2 - accuracy0 music phase3	0.0763	0.0447	2418.2399	1.7067	0.0880
accuracy0 control phase2 - accuracy1 music phase3	0.0964	0.0370	2414.5352	2.6076	0.0092
accuracy1 control phase2 - accuracy0 music phase2	0.0347	0.0347	2346.0755	1.0006	0.3171
accuracy1 control phase2 - accuracy1 music phase2	0.0029	0.0242	2404.7231	0.1189	0.9054
accuracy1 control phase2 - accuracy0 control phase3	-0.0541	0.0405	2404.8196	-1.3342	0.1823
accuracy1 control phase2 - accuracy1 control phase3	0.0177	0.0287	2404.3958	0.6174	0.5370
accuracy1 control phase2 - accuracy0 music phase3	-0.0067	0.0352	2298.7100	-0.1900	0.8493
accuracy1 control phase2 - accuracy1 music phase3	0.0134	0.0242	2405.9394	0.5558	0.5784
accuracy0 music phase2 - accuracy1 music phase2	-0.0318	0.0311	2308.1438	-1.0217	0.3070

accuracy0 music phase2 - accuracy0 control phase3	-0.0888	0.0447	2413.8733	-1.9844	0.0473
accuracy0 music phase2 - accuracy1 control phase3	-0.0170	0.0347	2353.6993	-0.4895	0.6246
accuracy0 music phase2 - accuracy0 music phase3	-0.0414	0.0399	2418.1131	-1.0373	0.2997
accuracy0 music phase2 - accuracy1 music phase3	-0.0213	0.0311	2325.5052	-0.6840	0.4940
accuracy1 music phase2 - accuracy0 control phase3	-0.0569	0.0375	2401.5384	-1.5169	0.1294
accuracy1 music phase2 - accuracy1 control phase3	0.0148	0.0243	2401.1911	0.6104	0.5417
accuracy1 music phase2 - accuracy0 music phase3	-0.0096	0.0317	2252.0253	-0.3017	0.7629
accuracy1 music phase2 - accuracy1 music phase3	0.0106	0.0188	2401.5038	0.5625	0.5738
accuracy0 control phase3 - accuracy1 control phase3	0.0718	0.0406	2395.9338	1.7675	0.0773
accuracy0 control phase3 - accuracy0 music phase3	0.0474	0.0450	2405.2354	1.0518	0.2930
accuracy0 control phase3 - accuracy1 music phase3	0.0675	0.0375	2397.6563	1.7988	0.0722
accuracy1 control phase3 - accuracy0 music phase3	-0.0244	0.0353	2271.2886	-0.6913	0.4895
accuracy1 control phase3 - accuracy1 music phase3	-0.0043	0.0242	2399.2263	-0.1762	0.8602
accuracy0 music phase3 - accuracy1 music phase3	0.0201	0.0317	2220.4392	0.6346	0.5257

REFERENCES

Amiez, C., & Petrides, M. (2007). Selective involvement of the mid-dorsolateral prefrontal cortex in the coding of the serial order of visual stimuli in working memory. *Proceedings of the National Academy of Sciences*, *104*(34), 13786–13791.

Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, *20*(2), 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7

Aron, A. R., Shohamy, D., Clark, J., Myers, C., Gluck, M. A., & Poldrack, R. A. (2004). Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of Neurophysiology*, *92*(2), 1144–1152.

Asaridou, S. S., Takashima, A., Dediu, D., Hagoort, P., & McQueen, J. M. (2016). Repetition Suppression in the Left Inferior Frontal Gyrus Predicts Tone Learning Performance. *Cerebral Cortex*, *26*(6), 2728–2742. https://doi.org/10.1093/cercor/bhv126

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, *21*(3), 170–176.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Barrientos, S. A., & Tiznado, V. (2016). Hippocampal CA1 Subregion as a Context Decoder. *Journal of Neuroscience*, *36*(25), 6602–6604. https://doi.org/10.1523/JNEUROSCI.1107-16.2016

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*, *83*, 62–78.

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Bellassen, V., Iglói, K., De Souza, L. C., Dubois, B., & Rondi-Reig, L. (2012). Temporal order memory assessed during spatiotemporal navigation as a behavioral cognitive marker for differential Alzheimer's disease diagnosis. *Journal of Neuroscience*, *32*(6), 1942–1952.

Berron, D., Schütze, H., Maass, A., Cardenas-Blanco, A., Kuijf, H. J., Kumaran, D., & Düzel, E. (2016). Strong Evidence for Pattern Separation in Human Dentate Gyrus. *Journal of Neuroscience*, *36*(29), 7569–7579. https://doi.org/10.1523/JNEUROSCI.0518-16.2016

Bialleck, K. A., Schaal, H.-P., Kranz, T. A., Fell, J., Elger, C. E., & Axmacher, N. (2011). Ventromedial Prefrontal Cortex Activation Is Associated with Memory Formation for Predictable Rewards. *PLOS ONE*, *6*(2), e16695. https://doi.org/10.1371/journal.pone.0016695

Blumenfeld, R. S., Parks, C. M., Yonelinas, A. P., & Ranganath, C. (2011). Putting the pieces together: The role of dorsolateral prefrontal cortex in relational memory encoding. *Journal of Cognitive Neuroscience*, *23*(1), 257–265.

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, *19*(7), 678–685.

Brown, T. I., Carr, V. A., LaRocque, K. F., Favila, S. E., Gordon, A. M., Bowles, B., Bailenson, J. N., & Wagner, A. D. (2016). Prospective representation of navigational goals in the human hippocampus. *Science*, *352*(6291), 1323–1326.

Brown, T. I., Hasselmo, M. E., & Stern, C. E. (2014). A High-resolution study of hippocampal and medial temporal lobe correlates of spatial context and prospective overlapping route memory. *Hippocampus*, *24*(7), 819–839. https://doi.org/10.1002/hipo.22273

Burunat, I., Alluri, V., Toiviainen, P., Numminen, J., & Brattico, E. (2014). Dynamics of brain activity underlying working memory for music in a naturalistic condition. *Cortex*, *57*, 254–269.

Burunat, I., Brattico, E., Hartmann, M., Vuust, P., Särkämö, T., & Toiviainen, P. (2018). Musical training predicts cerebello-hippocampal coupling during music listening. *Psychomusicology: Music, Mind, and Brain*, *28*(3), 152–152.

Burunat, I., Levitin, D. J., & Toiviainen, P. (2024). Breaking (musical) boundaries by investigating brain dynamics of event segmentation during real-life music-listening. *Proceedings of the National Academy of Sciences*, *121*(36), e2319459121. https://doi.org/10.1073/pnas.2319459121

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

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. https://doi.org/10.1111/j.1467-9280.2006.01801.x

Cox, W. R., Dobbelaar, S., Meeter, M., Kindt, M., & van Ast, V. A. (2021). Episodic memory enhancement versus impairment is determined by contextual similarity across events. *Proceedings of the National Academy of Sciences*, *118*(48), e2101509118–e2101509118.

Dagher, A., Owen, A. M., Boecker, H., & Brooks, D. J. (2001). The role of the striatum and hippocampus in planning: A PET activation study in Parkinson's disease. *Brain*, *124*(5), 1020–1032.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

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

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

DeVito, L. M., & Eichenbaum, H. (2011). Memory for the Order of Events in Specific Sequences: Contributions of the Hippocampus and Medial Prefrontal Cortex. *Journal of Neuroscience*, *31*(9), 3169–3175. https://doi.org/10.1523/JNEUROSCI.4202-10.2011

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

DuBrow, S., & Davachi, L. (2014). Temporal memory is shaped by encoding stability and intervening item reactivation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(42), 13998–14005. https://doi.org/10.1523/JNEUROSCI.2535-14.2014

DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of Learning and Memory*, *134*, 107–114.

Duncan, K., Ketz, N., Inati, S. J., & Davachi, L. (2012). Evidence for area CA1 as a match/mismatch detector: A high-resolution fMRI study of the human hippocampus. *Hippocampus*, *22*(3), 389–398. https://doi.org/10.1002/hipo.20933

Eichenbaum, H. (2013). Memory on time. *Trends in Cognitive Sciences*, *17*(2), 81–88. https://doi.org/10.1016/j.tics.2012.12.007

Eichenbaum, H. (2014). Time cells in the hippocampus: A new dimension for mapping memories. *Nature Reviews Neuroscience*, *15*(11), 732–744.

El Haj, M., Postal, V., & Allain, P. (2012). Music enhances autobiographical memory in mild Alzheimer's disease. *Educational Gerontology*, *38*(1), 30–41.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), Article 1. https://doi.org/10.1038/s41592-018-0235-4

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

Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron, 81*(5), 1179–1189.

Ezzyat, Y., & Davachi, L. (2021). Neural Evidence for Representational Persistence Within Events. *Journal of Neuroscience*, *41*(37), 7909–7920. https://doi.org/10.1523/JNEUROSCI.0073-21.2021

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Ferreri, L., Bigand, E., & Bugaiska, A. (2015). The positive effect of music on source memory. *Musicae Scientiae*, *19*(4), 402–411.

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), 458–458.

Franco, A., Cleeremans, A., & Destrebecqz, A. (2011). Statistical Learning of Two Artificial Languages Presented Successively: How Conscious? *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00229

Frank, L. E., Preston, A. R., & Zeithamova, D. (2019). Functional connectivity between memory and reward centers across task and rest track memory sensitivity to reward. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(3), 503–522. https://doi.org/10.3758/s13415-019-00700-8

Gerraty, R. T., Davidow, J. Y., Wimmer, G. E., Kahn, I., & Shohamy, D. (2014). Transfer of Learning Relates to Intrinsic Connectivity between Hippocampus, Ventromedial Prefrontal Cortex, and Large-Scale Networks. *Journal of Neuroscience*, *34*(34), 11297–11303. https://doi.org/10.1523/JNEUROSCI.0185-14.2014

Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013

Glicksohn, A., & Cohen, A. (2013). The role of cross-modal associations in statistical learning. *Psychonomic Bulletin & Review*, *20*, 1161–1169.

Gluth, S., Sommer, T., Rieskamp, J., & Büchel, C. (2015). Effective Connectivity between Hippocampus and Ventromedial Prefrontal Cortex Controls Preferential Choices from Memory. *Neuron*, *86*(4), 1078–1090. https://doi.org/10.1016/j.neuron.2015.04.023

Goodroe, S. C., Starnes, J., & Brown, T. I. (2018). The complex nature of hippocampal-striatal interactions in spatial navigation. *Frontiers in Human Neuroscience*, *12*, 250–250.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics*, *5*. https://doi.org/10.3389/fninf.2011.00013

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19*(5), 893–906.

Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: Premotor and striatal interactions in musicians and nonmusicians during beat perception. *Journal of Neuroscience*, *29*(23), 7540–7548.

Grahn, J. A., & Rowe, J. B. (2013). Finding and feeling the musical beat: Striatal dissociations between detection and prediction of regularity. *Cerebral Cortex*, *23*(4), 913–921.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Guzowski, J. F., Knierim, J. J., & Moser, E. I. (2004). Ensemble Dynamics of Hippocampal Regions CA3 and CA1. *Neuron*, *44*(4), 581–584. https://doi.org/10.1016/j.neuron.2004.11.003

Haber, S. N., & Knutson, B. (2010). The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology*, *35*(1), 4–26.

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

Hoge, J., & Kesner, R. P. (2007). Role of CA3 and CA1 subregions of the dorsal hippocampus on temporal processing of objects. *Neurobiology of Learning and Memory*, *88*(2), 225–231. https://doi.org/10.1016/j.nlm.2007.04.013

Howard, M. W., & Eichenbaum, H. (2015). Time and space in the hippocampus. *Brain Research*, *1621*, 345–354.

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Hsieh, L.-T., Ekstrom, A. D., & Ranganath, C. (2011). Neural Oscillations Associated with Item and Temporal Order Maintenance in Working Memory. *Journal of Neuroscience, 31*(30), 10803–10810. https://doi.org/10.1523/JNEUROSCI.0828-11.2011

Hsieh, L.-T., Gruber, M. J., Jenkins, L. J., & Ranganath, C. (2014). Hippocampal activity patterns carry information about objects in temporal context. *Neuron, 81*(5), 1165–1178.

Jarret, T., Stockert, A., Kotz, S. A., & Tillmann, B. (2019). Implicit learning of artificial grammatical structures after inferior frontal cortex lesions. *PloS One, 14*(9), e0222385–e0222385.

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage, 62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Johnson, D. L., & Kesner, R. P. (1997). Comparison of temporal order memory in early and middle stage Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 19*(1), 83–100.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science, 13*(2), 339–345.

Karuza, E. A., Emberson, L. L., Roser, M. E., Cole, D., Aslin, R. N., & Fiser, J. (2017). Neural Signatures of Spatial Statistical Learning: Characterizing the Extraction of Structure from Complex Visual Scenes. *Journal of Cognitive Neuroscience, 29*(12), 1963–1976. https://doi.org/10.1162/jocn_a_01182

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language, 127*(1), 46–54.

Kim, H., Wang, K., Cutting, L. E., Willcutt, E. G., Petrill, S. A., Leopold, D. R., Reineberg, A. E., Thompson, L. A., & Banich, M. T. (2022). The Angular Gyrus as a Hub for Modulation of Language-related Cortex by Distinct Prefrontal Executive Control Regions. *Journal of Cognitive Neuroscience*, *34*(12), 2275–2296. https://doi.org/10.1162/jocn_a_01915

Kim, S. M., Ganguli, S., & Frank, L. M. (2012). Spatial Information Outflow from the Hippocampal Circuit: Distributed Spatial Coding and Phase Precession in the Subiculum. *Journal of Neuroscience*, *32*(34), 11539–11558. https://doi.org/10.1523/JNEUROSCI.5942-11.2012

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Knutson, K. M., Wood, J. N., & Grafman, J. (2004). Brain activation in processing temporal sequence: An fMRI study. *NeuroImage*, *23*(4), 1299–1307. https://doi.org/10.1016/j.neuroimage.2004.08.012

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, *15*(3), 170–180.

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive Processes and the Peculiar Case of Music. *Trends in Cognitive Sciences*, *23*(1), 63–77. https://doi.org/10.1016/j.tics.2018.10.006

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79.

Leutgeb, J. K., Leutgeb, S., Moser, M.-B., & Moser, E. I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science*, *315*(5814), 961–966.

MacDonald, C. J., & Tonegawa, S. (2021). Crucial role for CA2 inputs in the sequential organization of CA1 time cells supporting memory. *Proceedings of the National Academy of Sciences*, *118*(3), e2020698118. https://doi.org/10.1073/pnas.2020698118

Matell, M. S., Meck, W. H., & Nicolelis, M. A. L. (2003). Interval timing and the encoding of signal duration by ensembles of cortical and striatal neurons. *Behavioral Neuroscience*, *117*(4), 760–760.

Maurer, A. P., & Nadel, L. (2021). The continuity of context: A role for the hippocampus. *Trends in Cognitive Sciences*, *25*(3), 187–199.

McAndrews, M. P., & Milner, B. (1991). The frontal cortex and memory for temporal order. *Neuropsychologia*, *29*(9), 849–859.

Mello, G. B. M., Soares, S., & Paton, J. J. (2015). A scalable population code for time in the striatum. *Current Biology*, *25*(9), 1113–1122.

Mitchel, A. D., & Weiss, D. J. (2011). Learning across senses: Cross-modal effects in multisensory statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1081–1081.

Mumford, J. A. (2013). Considerations when using Single-Trial parameter estimates in representational similarity analyses. *SHORT COURSE II*, 44.

Murray, L. J., & Ranganath, C. (2007). The dorsolateral prefrontal cortex contributes to successful relational memory encoding. *Journal of Neuroscience*, *27*(20), 5515–5522.

Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neuroscience & Biobehavioral Reviews*, *33*(7), 1004–1023. https://doi.org/10.1016/j.neubiorev.2009.04.001

Ninokura, Y., Mushiake, H., & Tanji, J. (2003). Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. *Journal of Neurophysiology*, *89*(5), 2868–2873.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Onuki, Y., Van Someren, E. J. W., De Zeeuw, C. I., & Van der Werf, Y. D. (2015). Hippocampal–cerebellar interaction during spatio-temporal prediction. *Cerebral Cortex*, *25*(2), 313–321.

Patriat, R., Reynolds, R. C., & Birn, R. M. (2017). An improved model of motion-related signal changes in fMRI. *NeuroImage*, *144*, 74–82. https://doi.org/10.1016/j.neuroimage.2016.08.051

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. https://doi.org/10.1016/j.bandl.2010.08.003

Pettijohn, K. A., Thompson, A. N., Tamplin, A. K., Krawietz, S. A., & Radvansky, G. A. (2016). Event boundaries and memory improvement. *Cognition*, *148*, 136–144.

Polti, I., Nau, M., Kaplan, R., van Wassenhove, V., & Doeller, C. F. (2022). Rapid encoding of task regularities in the human hippocampus guides sensorimotor timing. *Elife*, *11*, e79027–e79027.

Polyn, S. M., & Kahana, M. J. (2008). Memory search and the neural representation of context. *Trends in Cognitive Sciences*, *12*(1), 24–30. https://doi.org/10.1016/j.tics.2007.10.010

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). Task context and organization in free recall. *Neuropsychologia*, *47*(11), 2158–2163. https://doi.org/10.1016/j.neuropsychologia.2009.02.013

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Preston, A. R., & Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Current Biology*, *23*(17), R764–R773.

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences*, *17*, 133–140. https://doi.org/10.1016/j.cobeha.2017.08.006

Ranganath, C. (2010). Binding items and contexts: The cognitive neuroscience of episodic memory. *Current Directions in Psychological Science*, *19*(3), 131–137.

Ranganath, C., & Hsieh, L. (2016). The hippocampus: A special place for time. *Annals of the New York Academy of Sciences*, *1369*(1), 93–110.

Ren, Y., Leslie, G., & Brown, T. (2024). Visual sequence encoding is modulated by music schematic structure and familiarity. *PLOS ONE*, *19*(8), e0306271. https://doi.org/10.1371/journal.pone.0306271

Ren, Y., Mehdizadeh, S. K., Leslie, G., & Brown, T. (2024). Affective music during episodic memory recollection modulates subsequent false emotional memory traces: An fMRI study. *Cognitive, Affective, & Behavioral Neuroscience*, *24*(5), 912–930. https://doi.org/10.3758/s13415-024-01200-0

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.

Rolls, E. (2013). A quantitative theory of the functions of the hippocampal CA3 network in memory. *Frontiers in Cellular Neuroscience*, *7*. https://doi.org/10.3389/fncel.2013.00098

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

Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2020). Reward prediction errors create event boundaries in memory. *Cognition*, *203*, 104269–104269.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*, *14*(2), 257–262.

Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., & Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science, 340*(6129), 216–219.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Elger, C. E., Friederici, A. D., Grigutsch, M., Huppertz, H., Knösche, T. R., & Wellmer, J. (2009). Overlap of musical and linguistic syntax processing: Intracranial ERP evidence. *Annals of the New York Academy of Sciences*, *1169*(1), 494–498.

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

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., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, *22*(17), 1622–1627.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486–492. https://doi.org/10.1038/nn.3331

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, *26*(1), 3–8.

Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, *37*(6), 1013–1025.

Schmidt, B., Marrone, D. F., & Markus, E. J. (2012). Disambiguating the similar: The dentate gyrus and pattern separation. *Behavioural Brain Research*, *226*(1), 56–65. https://doi.org/10.1016/j.bbr.2011.08.039

Sherman, B. E., Graves, K. N., & Turk-Browne, N. B. (2020). The prevalence and importance of statistical learning in human cognition and behavior. *Current Opinion in Behavioral Sciences*, *32*, 15–20.

Skosnik, P. D., Mirza, F., Gitelman, D. R., Parrish, T. B., Mesulam, M.-M., & Reber, P. J. (2002). Neural Correlates of Artificial Grammar Learning. *NeuroImage*, *17*(3), 1306–1314. https://doi.org/10.1006/nimg.2002.1291

Smith, S. M. (1985). Background Music and Context-Dependent Memory. *The American Journal of Psychology*, *98*(4), 591–591. https://doi.org/10.2307/1422512

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*(2), 203–220. https://doi.org/10.3758/BF03196157

Sridharan, D., Levitin, D. J., Chafe, C. H., Berger, J., & Menon, V. (2007). Neural Dynamics of Event Segmentation in Music: Converging Evidence for Dissociable Ventral and Dorsal Networks. *Neuron*, *55*(3), 521–532. https://doi.org/10.1016/j.neuron.2007.07.003

St. Jacques, P., Rubin, D. C., LaBar, K. S., & Cabeza, R. (2008). The short and long of it: Neural correlates of temporal-order memory for autobiographical events. *Journal of Cognitive Neuroscience*, *20*(7), 1327–1341.

Sullivan, E. V., & Sagar, H. J. (1989). Nonverbal recognition and recency discrimination deficits in Parkinson's disease and Alzheimer's disease. *Brain*, *112*(6), 1503–1517.

Sun, Y., Jin, S., Lin, X., Chen, L., Qiao, X., Jiang, L., Zhou, P., Johnston, K. G., Golshani, P., Nie, Q., Holmes, T. C., Nitz, D. A., & Xu, X. (2019). CA1-projecting subiculum neurons facilitate object–place learning. *Nature Neuroscience*, *22*(11), 1857–1870. https://doi.org/10.1038/s41593-019-0496-y

Tillmann, B., & McAdams, S. (2004). Implicit Learning of Musical Timbre Sequences: Statistical Regularities Confronted With Acoustical (Dis)Similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1131–1142. https://doi.org/10.1037/0278-7393.30.5.1131

Tulving, E. (1983). *Elements of episodic memory*.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*(1), 1–25.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–552.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945. Journal of Cognitive Neuroscience. https://doi.org/10.1162/jocn.2009.21131

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, *30*(33), 11177–11187.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

van Buuren, M., Kroes, M. C. W., Wagner, I. C., Genzel, L., Morris, R. G. M., & Fernández, G. (2014). Initial investigation of the effects of an experimentally learned schema on spatial associative memory in humans. *Journal of Neuroscience*, *34*(50), 16662–16670.

van der Meer, M. A. A., Johnson, A., Schmitzer-Torbert, N. C., & Redish, A. D. (2010). Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. *Neuron*, *67*(1), 25–32.

van Kesteren, M. T. R., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. *Neuropsychologia*, *51*(12), 2352–2359. https://doi.org/10.1016/j.neuropsychologia.2013.05.027

van Kesteren, M. T. R., Fernández, G., Norris, D. G., & Hermans, E. J. (2010). Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and

postencoding rest in humans. *Proceedings of the National Academy of Sciences*, *107*(16), 7550–7555. https://doi.org/10.1073/pnas.0914892107

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

Van Opstal, F., Verguts, T., Orban, G. A., & Fias, W. (2008). A hippocampal–parietal network for learning an ordered sequence. *NeuroImage*, *40*(1), 333–341.

Wallace, W. T. (1994). Memory for music: Effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1471–1471.

Wallenstein, G. V., Hasselmo, M. E., & Eichenbaum, H. (1998). The hippocampus as an associator of discontiguous events. *Trends in Neurosciences*, *21*(8), 317–323.

Wanjia, G., Favila, S. E., Kim, G., Molitor, R. J., & Kuhl, B. A. (2021). Abrupt hippocampal remapping signals resolution of memory interference. *Nature Communications*, *12*(1), 4816. https://doi.org/10.1038/s41467-021-25126-0

Watkins, M. J., & Tulving, E. (1975). Episodic memory: When recognition fails. *Journal of Experimental Psychology: General*, *104*(1), 5–5.

Wen, T., & Egner, T. (2022). Retrieval context determines whether event boundaries impair or enhance temporal order memory. *Cognition*, *225*, 105145–105145.

Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn: A functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connectivity*, *2*(3), 125–141. https://doi.org/10.1089/brain.2012.0073

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, *21*(4), 1732–1747. https://doi.org/10.1016/j.neuroimage.2003.12.023

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006

Yildirim, I., & Jacobs, R. A. (2015). Learning multisensory representations for auditory-visual transfer of sequence category knowledge: A probabilistic language of thought approach. *Psychonomic Bulletin & Review*, *22*, 673–686.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, *31*(3), 1116–1128. https://doi.org/10.1016/j.neuroimage.2006.01.015

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction Error Associated with the Perceptual Segmentation of Naturalistic Events. *Journal of Cognitive Neuroscience*, *23*(12), 4057–4066. https://doi.org/10.1162/jocn_a_00078

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29–58. https://doi.org/10.1037/0096-3445.130.1.29

Zhang, D., Zhang, X., Sun, X., Li, Z., Wang, Z., He, S., & Hu, X. (2004). Cross-modal temporal order memory for auditory digits and visual locations: An fMRI study. *Human Brain Mapping*, *22*(4), 280–289.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. IEEE Transactions on Medical Imaging. https://doi.org/10.1109/42.906424