HELSINGIN YLIOPISTO

The hippocampal audio editor

Neural correlates of event segmentation in an auditory narrative

Cognitive science / Cognitive neuroscience
Master's Thesis

Author:
Jenni Saaristo

Supervisors:
Prof. Iiro Jääskeläinen
Prof. Riikka Möttönen

18.1.2023
Helsinki

Tiivistelmä

Tiedekunta: Humanistinen tiedekunta

Koulutusohjelma: Kognitiotiede

Opintosuunta: Kognitiivinen neurotiede

Tekijä: Jenni Saaristo

Työn nimi: Tapahtumasegmentaation aivovasteet hippokampuksessa ja aivokuorella äänitarinan kuuntelun aikana

Työn laji: Maisterintutkielma

Kuukausi ja vuosi: Tammikuu 2023

Sivumäärä: 77

Avainsanat: tapahtumakognitio, tapahtumasegmentaatio, fMRI, funktionaalinen magneettikuvaus, hippokampus, aivoturso, luonnolliset ärsykkeet

Ohjaajat: Prof. Iiro Jääskeläinen, Prof. Riikka Möttönen

Säilytyspaikka: Helsingin yliopiston kirjasto

Muita tietoja:

Tiivistelmä: Tapahtumasegmentaatio jäsentää sekä arkista kokemustamme että muistiamme. Parhaillaan meneillään olevan tapahtuman hahmotus ja prosessointi tapahtuu todennäköisesti aivokuorella, mutta ilman toimivaa hippokampusta tilanteesta ei voi syntyä pysyvää muistoa. On olennainen kysymys, missä kohtaa ja miten hippokampus osallistuu tapahtumien prosessointiin ja mieleen painamiseen. Aiemmin on magneettikuvaustutkimuksin osoitettu, että hippokampus reagoi tapahtumien välisiin rajoihin aktivaatiopiikein. On ehdotettu, että ne ilmentäisivät aistimodaliteetista riippumattoman tason prosessia, jossa hippokampus kokoaa yhteen ja vahvistaa koetun tilanteen kokonaisrepresentaation, jotta se voidaan painaa muistiin. Aiemmat tutkimukset on kuitenkin toteutettu yksinomaan audiovisuaalisilla ärsykkeillä, ja koska hippokampuksen tiedetään osallistuvan myös visuaaliseen prosessointiin, ei ole täysin selvää, etteivätkö havaitut aktivaatiot voisi selittyä alemman, aistitietoa käsittelevän tason prosesseilla.

Tämän kysymyksen ratkaisemiseksi tässä tutkimuksessa selvitettiin reagoiko hippokampus tapahtumarajoihin puhtaasti auditiivisessa ärsykkeessä. Ärsykkeenä oli 71-minuuttinen tarinallinen äänikirja, jonka osallistujat kuuntelivat passiivisesti fMRI-rekisteröinnin aikana, ja jonka tapahtumarajat määriteltiin kokeellisesti erillisen koehenkilöryhmän avulla. Aivokuvausaineisto analysoitiin aivoalueittain sekä hippokampuksesta että eksploratiivisesti myös kaikilta aivokuoren alueilta.

Hippokampuksen havaittiin reagoivan tapahtumarajoihin aktivaatiopiikein. Aivokuorella voimakkaasti reagoivia alueita olivat mm. posteriorinen mediaalinen aivokuori, ventromediaalinen prefrontaalialue, parahippokampaalinen poimu sekä etummainen pihtipoimu. Monien näistä alueista uskotaan osallistuvan meneillään olevan tapahtuman mallintamiseen ja hahmottamiseen, ja osa mahdollisesti osallistuu huomion siirtämiseen sisäisen ja ulkoisen välillä. Etummaisen pihtipoimun tiedetään osallistuvan odotusten ja havaintojen välisten konfliktien monitorointiin, mikä saattaisi tukea teoriaa, jonka mukaan segmentaatio olisi riippuvaista havaituista ennustevirheistä. Tätä ei kuitenkaan tämän tutkimuksen perusteella voida varmasti päätellä, vaan asiaa tulisi tutkia tarkemmin.

Tämän tutkimuksen tulokset tukevat näkemystä, jonka mukaan hippokampuksen lisääntynyt toiminta tapahtumarajoilla liittyy korkean tason abstraktiin segmentaatioon ja mahdollisesti episodisen muiston luomiseen. Tämä prosessi mahdollisesti tapahtuu yhteistyössä aivokuoren aktiivisten alueiden kanssa, mutta kausaaliset suhteet ja informaation kulku näiden alueiden välillä on selvitettävä myöhemmissä tutkimuksissa.

Abstract

Faculty: Faculty of Arts

Degree programme: Cognitive science

Study track: Cognitive neuroscience

Author: Jenni Saaristo

Title: The hippocampal audio editor – neural correlates of event segmentation in an auditory narrative

Level: Master's thesis

Month and year: Jan 2023

Number of pages: 77

Keywords: event cognition, event segmentation, event boundaries, fMRI, hippocampus, naturalistic stimuli

Supervisors: Prof. Iiro Jääskeläinen, Prof. Riikka Möttönen

Where deposited: Helsinki University Library

Additional information:

Abstract: Event segmentation structures our experience as well as our memories. The representation of the currently ongoing event is likely dependent on a network of cortical areas, but the ability to retain a memory of the event requires an intact hippocampus. It is thus a relevant question how and when this hippocampal episodic encoding happens. It has previously been shown that the hippocampus is sensitive to event boundaries and responds to them with transient fMRI activation peaks. It has been proposed that these hippocampal end-of-event activations represent a high-level, modality-independent process of sharpening or “printing out” of the memory trace of the situation. However, the studies reporting hippocampal peaks have been conducted on audio-visual stimuli, so it is unclear whether these results generalise to narratives without a visual component, as the hippocampus is known to support visual processing as well as episodic encoding.

In this study I aim to answer this question by analysing fMRI data from participants experiencing a purely auditory narrative. The stimulus was a 71-minute-long audio book, and it was segmented behaviourally by a separate group of participants with a naïve intuitive segmentation paradigm. The data was analysed with a region of interest (ROI) analysis in the hippocampus, as well as in an exploratory manner on all areas from a cortical atlas.

The hippocampus was found to respond significantly to event boundaries in the story. Strong responses were also found in areas of the posterior medial cortex (PMC), as well as in ventromedial prefrontal cortex (vmPFC), parahippocampal gyrus, anterior cingulate (ACC) and the insula. Many of these are known to be involved in representing the event model, and some with switching between internal and external processing modes. ACC in particular is known to be involved in conflict monitoring – this might link with the proposal that segmentation in general is driven by prediction error and would merit further study.

I conclude that the hippocampus does detect and respond to event boundaries in a naturalistic auditory narrative, which is in line with the “print out” hypothesis and implies that these activations are related to domain-general episodic encoding. The increased hippocampal processing is likely to happen in collaboration with cortical areas involved in signalling change and representing the working event model. However, the causal connections between these areas during the boundary-related processing cascade needs to be elaborated in future studies.

Preface

A functional brain imaging study of this scale is not a one-person show. All along the way I have received help and support from the Aalto neuroimaging community at the Department of Neuroscience and Biomedical Engineering, especially in the Brain & Mind Laboratory, headed by Professor Iiro Jääskeläinen. Using pre-existing neuroimaging data from a study by Maria Hakonen enabled me to concentrate fully on the analysis phase. Emma Stickler started this particular project before I came to it, set up the behavioural experiment and collected some of the data.

This study is in many ways an attempt to replicate of the findings of Aya Ben-Yakov and Rik Henson (published in 2018), and I received much methodological insights from comments and discussions with Aya, of which I am grateful. Maria Hakonen helped with the data and redid the preprocessing for my benefit, as well as offered many a good advice and encouragement on the way. I also gratefully acknowledge the computational resources provided by the Aalto Science-IT project and especially the technical and methodological support it came with. Iiro Jääskeläinen enabled me to work on this project which, while daunting, turned out to suit me perfectly, and he kept supporting me, though the process turned lengthy. Riikka Möttönen asked pointed questions and gently but unyieldingly made me strive for deeper understanding.

In addition to these I would like to thank Mareike Bacha-Trams for helping me to keep my head together, Liisa Rantalaiho for persistently enquiring after the state of this thesis and providing support and comments (as well as generally being an inspiration), and my family, friends, and neighbours for helping me tide through troughs of desperation.

And, lastly, I would like to thank all the unfortunate people who chanced to ask, “So what’s your thesis about?” and launched me into a tirade that somehow, finally, became the backbone of this thesis.

Contents

1 Introduction

Despite the continuity of our stream of experience, people naturally and automatically segment ongoing activity into discrete events or episodes (Newtson, 1973; Zacks et al., 2001). Whether we describe our day, or a YouTube video we just saw, we are drawn to reporting it in terms of events and their "gist" rather than the detailed perceptual impressions they contained. When asked to "segment" or place boundaries between events in many types of temporally dynamic stimuli people tend to agree on their location (Sasmita & Swallow, 2021; Speer et al., 2003). It has even been proposed that the mechanisms of event perception are fundamental organisers of our conscious experience (Zacks, 2020).

The classical definition of an event in the event segmentation literature is "a segment of time that is conceived by an observer to have a beginning and an end" (Zacks et al., 2001). In general parlance an event is often thought of as something that might represent an episode of episodic memory – e.g. attending a wedding ceremony, or having dinner with friends – or perhaps a scene in a film, signified by continuity in time, place and characters. These types of episodes typically span from some tens of seconds to several hours.

The phenomenon of segmentation has connections with two fundamental principles of human cognition: predictive encoding and episodic memory. The latter has been proposed to be the eminently human ability to revisit the experiences of the past (Tulving, 1972, 2002). The subject matter of this mental time travel is an episode: an integrated representation containing the subjectively essential features of the experience, connected both relationally and sequentially, with one part of the memory obligatorily triggering many others and organised in a temporal chronology.

Episodic memory is heavily dependent on the hippocampus (HC). A damage to the medial temporal lobe (MTL) – which consists of HC and connected structures – typically leads to amnesia, which is a specific inability to establish new declarative memories despite mostly intact intellectual, social and perceptual capacities (reviewed in, e.g., Squire et al., 2004; Squire & Zola-Morgan, 1988; Tulving, 2002).

It is believed that especially HC is critical for one-shot learning, which is required for both episodic and semantic long-term memory – without it, while new information and skills may be gradually learned through numerous repetitions, there is no conscious recollection of learning, and the learned material cannot be explicitly articulated (Squire & Zola-Morgan, 1988; Zola-Morgan et al., 1986).

The other major principle connected with event segmentation is predictive encoding: most of our mnemonic capacity expressly serves the future – essentially, we learn in order to predict. Segmenting of experiences, while subserving their recollection, also serves our ability to predict both the general arc of a situation as well as the items that are likely to feature in it. To enable this the internal event model has to capture the statistics of a situation accurately and "carve activity at its joints", i.e., separate events that are not connected in a way that would be useful for predictions (Richmond & Zacks, 2017). These points of discontinuity and distinctive change are proposed to give rise to event boundaries (Newtson et al., 1977; Zacks et al., 2007).

Event boundaries were originally described and studied as junction points in goal-directed physical movements by human actors (Newtson, 1973; Speer et al., 2003; Zacks et al., 2001) but increasingly the research has concentrated on narrative stimuli. This seems like a logical development, given how narrative structure mirrors our understanding of time, events and causality, and capitalises on that to deliver information in a highly impactful and memorable way. Thus, while research on all possible levels of segmentation is worthwhile, narrative level promises most gains in elucidating how we process and remember events, and use them socially to communicate and learn about the world and ourselves (Lee et al., 2020). Naturalistic stimuli are seldom easy to work with, but the advances in understanding they have delivered are substantial (Jääskeläinen et al., 2021; Sonkusare et al., 2019), and they can help bridge the gap between cognition-in-the-lab and cognition-in-the-wild (Nastase et al., 2020).

In this thesis I will report findings from a study investigating the hippocampal responses to event boundaries in a naturalistic auditory narrative. Event boundaries in extended dynamic stimuli have been found to elicit transient increases in the blood oxygen level dependent (BOLD) signal in many cortical brain areas (reviewed in Richmond & Zacks, 2017), and they also have behavioural and memory consequences (reviewed in Chris M Bird, 2020). It is thus a reasonable assumption that also the hippocampus is involved in boundary processing, and some evidence of this exists from the audio-visual domain (e.g. Ben-Yakov & Dudai, 2011; Ben-Yakov & Henson, 2018). This study targets the hippocampus specifically, though I also investigate cortical regions-of-interest (ROIs).

Nearly all previous research on either event boundaries or hippocampal activations in episodic memory encoding have been conducted on visual stimuli: either static or dynamic, and with or without audio. Testing hypotheses on a purely auditory stimulus is an important step towards a more general understanding of boundary-related neural processing, as well as for assessing the domain-generality of the hippocampal memory-related response.

First I will introduce the concept and theories of event segmentation, as well as survey the current state of functional magnetic resonance imaging (fMRI) research done on event boundaries. Majority of the preceding literature on event boundaries has been done with fMRI, and as it is also the modality of our data, I will restrict myself to the systematic review of those studies. Then I will shortly describe what is known about the connections between hippocampus and episodic memory, and how that picture is changing with increased use of naturalistic stimuli in memory research. Finally, I will present the current study and its findings, and discuss their meaning in terms of event segmentation and memory formation.

1.1 What is event segmentation

Event Segmentation Theory (EST) was proposed as a framework to describe the phenomenon of segmentation and event perception in general (Zacks et al., 2007), and it maintained that event segmentation is automatic, multimodal, and hierarchical, as well as tightly linked to long-term memory (LTM). The latter means, for example, that segmentation is affected by prior knowledge (encapsulated in e.g. event schemata) and in turn affects how experiences are recorded in LTM. EST also asserts that segmentation during an experience is driven by prediction error, i.e., the difference between a predicted outcome and perceptual input. Many of these propositions have received abundant empirical support over the years, while others are somewhat contested. In this section I will examine some evidence for the tenets of EST, introduce some further concepts central to event cognition, and finally briefly delineate some unresolved questions.

Event segmentation as a phenomenon is thought to be an automatic and obligatory part of the processing of temporal information. It was first found out by Darren Newtson (1973) that if people were shown videos of actors performing everyday activities, and asked to identify boundaries between meaningful action units, they tended to agree on the location of those boundaries. Later, with the advent of non-invasive functional brain imaging, it was shown that these points of segmentation elicit time-locked brain activations – even when people are merely passively watching the stimulus and given no task at all (Zacks et al., 2001).

Since then, the assessment of exactly which brain areas are important might have changed – I will return to this later in the review of relevant fMRI literature – but the essential phenomenon has not: study after study has shown robust time-locked activations in response to event boundaries defined either afterwards by the same participants (e.g. Speer et al., 2003, 2007; Zacks et al., 2001, 2006, 2010) or by an entirely separate group of annotators (Ben-Yakov & Henson, 2018), underlining the automatic nature of boundary processing. Further, these studies have utilised a wide variety of stimuli, from everyday activities and extremely simple animations to narrative texts and full-length feature films, highlighting the multimodality of the phenomenon.

Segmentation is thought to be hierarchical, in the sense that people may simultaneously segment on several timescales or “grains”, and the “fine” boundaries are nested within those defined with a “coarse” grain (Newtson, 1973; Zacks et al., 2001, 2009, 2010). While the nesting effect has been shown in a number of studies, the argument is somewhat weakened by the fact that in most studies the subjects saw the stimuli several times (e.g., Zacks et al., 2001, 2009, 2010), which is likely to lead to learning effects; some studies even explicitly trained the subjects on a practice stimulus to segment at the correct grain before the actual task (Zacks et al., 2009, 2010), which weakens the argument for a naturally occurring tendency to segment at these specific levels. Regardless, the argument for hierarchical structuring seems generally plausible and has been widely accepted, and recent data-driven imaging studies give reason to expect that a more detailed view of the temporal hierarchy of event processing is about to emerge (e.g., Baldassano et al., 2017; Hasson et al., 2015).

A key tenet of EST is the argument that segmentation is driven by prediction error: the goal of event segmentation, according to the theory, is to create relatively stable event models which facilitate the prediction of upcoming activity (Zacks et al., 2007). When those predictions fail, it can be deduced that the situation has changed enough for the current model to have become useless. This would then trigger an event boundary, leading to the disposal of the current model and the construction of a new one (see Figure 1). This in turn would lead to two kinds of memory effects: firstly, as the current model is abandoned, the information comprising it is lost unless it is encoded into LTM – and even if it is encoded, it takes on a different form with much loss of perceptual detail. Secondly, as the construction of the new model requires paying more attention to incoming sensory stimuli, the items present at that time receive more thorough processing and are subsequently better remembered – an effect known as boundary advantage.

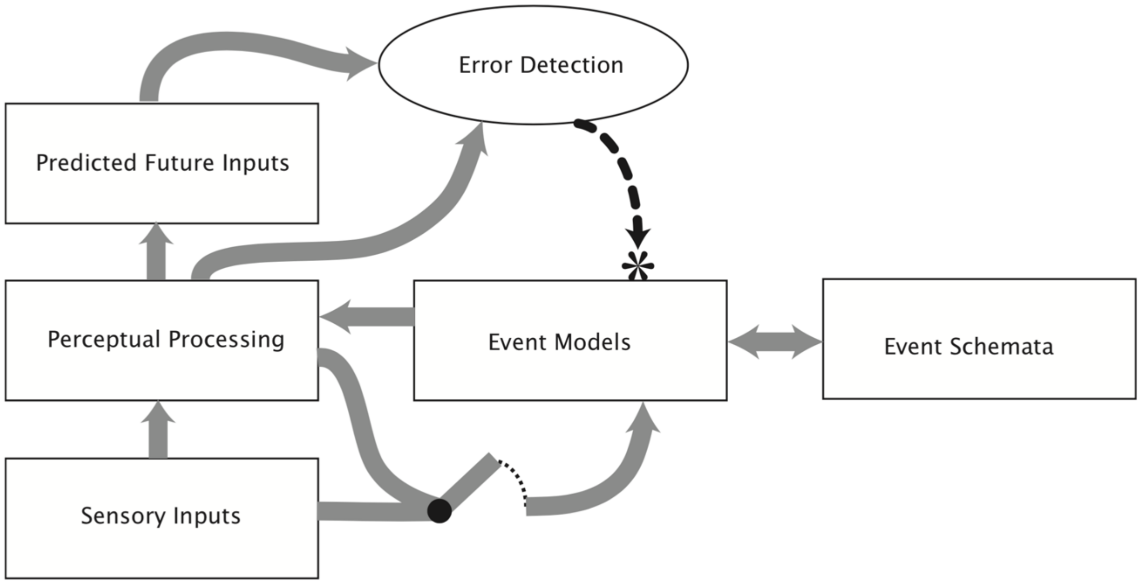


Figure 1: The original Event Segmentation Theory (from Zacks et al., 2007). A prediction error, i.e., a large enough difference between predicted and actual perceptual inputs triggers an event boundary. This leads to the updating of the current event model, with increased input of sensory information.

The evidence for boundary advantage is somewhat mixed: it has been found in some studies (Huff et al., 2018; Newtson & Engquist, 1976; Swallow et al., 2009), while others have reported the opposite, i.e., attention to visual probes is diminished if they coincide with an event boundary (Huff et al., 2012). This discrepancy is possibly due to differences in stimuli, boundaries, and other details in experimental design, the review of which is outside the scope of this thesis. Yet, it can be said that the question of attention deployment at event boundaries is still a rather open one. On the other hand, the memory consequences following from crossing an event boundary and discarding the working event model are rather striking and are explained in more detail below.

1.1.1 Event models shape memory and cognition

Event model is a key concept in event cognition, and before going into the effects it has on memory performance, we will briefly examine what is usually meant by it.

To explain event models it is useful to first introduce the concept of situation models used in the realm of language research. In 1995 Zwaan and others proposed that people code narrative texts as situation models, i.e. representations encompassing time, space, actors and their goals, and causal links between actions (Zwaan et al., 1995; Zwaan & Radvansky, 1998). Similarly, an event model represents aspects of the situation which are likely to remain relatively stable for a period of time and thus facilitate useful predictions (Richmond & Zacks, 2017; Zacks et al., 2007). While situation model refers to language-based, abstract understanding, event models also encompass multimodal representations of the objects, people, locations and actions involved, e.g. what they look like, how they sound, or how they are likely to move.

Event models are working memory (WM) representations – unlike event schemata, which are LTM representations of typical event structures and their associated items, event models contain information pertaining to the *specific* event, with the features that differentiate it from all other similar experiences (Zacks et al., 2007). Sometimes to emphasise this link between working memory and event models, the term working event model is used (Zacks, 2020). Event models are closely related to the episodic buffer – a conceptual WM space capable of maintaining an “expanded present” without creating any persistent memory traces (Baddeley, 2000; Baddeley & Wilson, 2002). It could be said that, theoretically, event models are the representations that populate the episodic buffer.

There are a number of findings to indicate that event models are built and discarded at event boundaries, leading to decreased subsequent memory across boundaries. It has been shown that crossing event boundaries leads to surface information loss, e.g. diminished recognition of visual or textual details (Gernsbacher, 1985; Speer & Zacks, 2005; Swallow et al., 2009, 2011). Also, doorways seem to elicit forgetting: Radvansky and colleagues have put participants through experiments which require remembering associated ("currently carried") and dissociated ("recently put down") abstract objects, while moving through a sequence of rooms. They have consistently found that experiencing a spatial shift decreases memory performance – especially for associated objects – whether the experiments are run in virtual spaces (Pettijohn & Radvansky, 2018; Radvansky & Copeland, 2006; Seel et al., 2019), a real space (Radvansky et al., 2011), or an imagined one (Lawrence & Peterson, 2016). Further, memory is not merely tied to context: returning to the same room after a shift does not help memory performance (Radvansky et al., 2011).

Note, though, that many of these experiments used changes in some dimension of a situation model – e.g. time or location – to define a boundary, and did not directly test whether or not they were perceived as event boundaries. Yet, when these situational changes (often called narrative shifts) have been investigated concurrently with subjective event boundaries in other studies, the two have been found to correlate significantly (Huff et al., 2014, 2018; Speer et al., 2007; Zacks et al., 2009). It is thus reasonable to assume that the effects discovered for narrative shifts mostly apply for event boundaries as well.

Another line of evidence concerns increased processing times at boundaries. This effect has been found in some of the studies on doorways described above, wherein probes following a spatial shift have elicited longer reaction times than those delivered in the same space with the object (Seel et al., 2019), while others have not found a significant effect, or the results have been mixed (Radvansky et al., 2011; Radvansky & Copeland, 2006). Also studies with verbal stimuli have reported rather mixed results, with some types of shifts consistently eliciting increased reading times and others not (reviewed in Pettijohn & Radvansky, 2016b). It is thus not altogether clear whether this line of evidence in fact supports the proposal that constructing a new event model or reaching across an event boundary requires extended processing and thus more time.

The third and most recent line of enquiry into event models involves functional brain imaging, and the finding that some high-level brain areas seem to track and segment events in a way that corresponds with the behavioural data. In 2017 Baldassano and others published a study on a fMRI dataset collected while the participants viewed the first 50 minutes of the opening episode of the BBC series Sherlock. The authors had developed a method of segmenting brain activity into “neural events” in a data-driven manner, based on detecting episodes of relatively stable activity patterns punctuated by rapid shifts (Baldassano et al., 2017). They found a hierarchy of neural segmentation, and at its highest level – e.g., in the posterior medial cortex (PMC) and the angular gyrus (AG) – the neural event boundaries coincided significantly with behaviourally defined event boundaries. This was a striking demonstration of segmentation as a naturally occurring phenomenon in the brain, as well as an indication that event models might be encoded as patterns of activation in certain areas of the cortex.

Two other studies investigated the content of these cortical event patterns, though with boundaries defined by major narrative shifts in the film and not the neural data itself. Chen and others reported that activity patterns in many high-order areas, including the PMC, were reinstated during free recall, and were shared across participants (Chen et al., 2017). Zadbood and others then showed that the same patterns were also duplicated in naïve participants who only experienced a recall-version of the story, and the strength of the neural alignment correlated with the listener’s level of comprehension (Zadbood et al., 2017). In other words, these event patterns seem to capture the essential dimensions of a situation and form the content that is transmitted in verbal communication.

These and other findings have led to the proposition that the major midline default mode network (DMN) hubs – the medial prefrontal cortex (mPFC) and the posterior cingulate cortex (PCC) – implement the functionality of the working event model (Stawarczyk et al., 2021). Also, it is possible that at least some portion of the mPFC is specifically tuned to event schemata, whereas PCC seems to be mostly agnostic in terms of prototypical scripts (Baldassano et al., 2018).

In summary, the working event model is proposed as a theoretical construct that integrates the elements of the ongoing situation, as well as connects with relevant information in the LTM. It bears similarities to Baddeley's episodic buffer but is more detailed in terms of both mechanisms and function. Event models seem to be discarded or updated at event boundaries, which makes information on the previous event less accessible, and in that sense they shape the structure of memory. Event models themselves do not seem to require the capability of long-term memory encoding, and they might explain the spared immediate prose recall abilities of some hippocampal amnesiacs (see e.g. Baddeley & Wilson, 2002). Retaining information across an event boundary likely requires it to be encoded into LTM, but to my knowledge this has not been directly tested with amnesic patients – in the delayed prose recall test the recall is delayed 20 minutes, with no regard to possible event boundaries. Finally, neural activation patterns in some areas of the cortex – namely PCC, vmPFC, and possibly AG – seem to represent events, as the rhythm of their shifts tends to coincide with subjective event boundaries and reporting the gist of events verbally to others causes corresponding patterns to emerge in their brains as well.

1.1.2 Is segmentation driven by prediction error?

EST proposed a somewhat mechanistical model for event segmentation, which relied heavily on prediction error: the ending of the ongoing event would elicit a failure in predicting the perceptual input, which would trigger segmentation and thus an event model update (see Figure 1). Recently this proposal has met with some criticism.

In the original EST model segmentation is *entirely* dependent on something unpredicted happening: as long as the predictions hold, nothing needs to be changed. Thus an expected sequence of daily events – such as waking up, going to school, attending the classes you should be attending, going home for a snack, then for the sports training, etc. – would demand no updates whatsoever; yet we know that these changes in location and activity are almost certain to trigger event boundaries (e.g. Magliano et al., 2014; Pettijohn & Radvansky, 2016a).

The most direct evidence in favour of prediction error as a trigger for segmentation is an fMRI study by Zacks and colleagues in 2011, which explicitly asked people to predict the continuation of a video stimulus within and across event boundaries (Zacks et al., 2011). The videos depicted everyday activities (e.g., washing a car), which naturally consist of several subunits of activity, but were unedited and otherwise un-manipulated. About once a minute the videos were paused, and the participants were asked in a two-alternative-forced-choice paradigm to predict which of two still images was likely to occur after 5 seconds – a delay which either contained or did not contain a behaviourally defined event boundary. They found that prediction accuracy and confidence was lower across event boundaries, and that the attempt to predict in that condition was accompanied with increased BOLD activity in the right substantia nigra, known to signal prediction error via dopaminergic pathways.

It is unclear how strong this evidence is for backing up the proposal for a general role of prediction error in event segmentation. While the behavioural effect is clear, it is limited by the nature of the stimulus, which in some sense consists of only one event (e.g., "doing the dishes"). Segmentation in this case thus pertains to smaller units than, for example, in a narrative stimulus. Further, the decrease in predictability across a boundary does not require that the boundary itself is defined by prediction error, and dopaminergic signalling when engaged in predictive decision making does not equal signalling at an event boundary.

A more direct way of showing triggering by prediction error would be, for instance, to detect prediction error related neural responses in passive viewing at naturalistic event boundaries. In fact, when introducing EST in 2007 Zacks and colleagues predicted that error signals should be observed as peaks in the anterior cingulate cortex (ACC) at event boundaries (Zacks et al., 2007). So far no study has found such evidence. In later updates of EST the hypotheses concerning ACC and MDS have largely been abandoned (e.g. Richmond & Zacks, 2017; Zacks, 2020).

Generally, the level of unexpectedness of event boundaries has not in the literature been tested, but rather assumed to be high. An exception to this is a study by Pettijohn and Radvansky, which investigated reading times and boundary detection in short written stories (Pettijohn & Radvansky, 2016b). The stimuli were explicitly manipulated in terms of the expectedness of the event shifts by means of foreshadowing, e.g. sentences giving a hint that the event is about to end without revealing what the exact change will be. The authors found that while foreshadowing decreased the unexpectedness of the shifts, it did not affect boundary detection. In other words, unpredictability did not play a role in determining the event boundaries. Further, they found that the same memory effects (i.e., reduced recall of event details) followed both types of shifts, indicating that an event model updating had taken place regardless of unpredictability.

In 2017 Richmond and Zacks published an updated view on EST and discussed some other possible triggering rules for segmentation (Richmond & Zacks, 2017). They concurred that instead of a point prediction the event modelling system may generate a distribution of predictions, and if their range becomes sufficiently large that would spell a need to segment. In other words, the future has become *uncertain*, and it is predicted that any prediction made now is likely to fail. The proposition is elegant and compelling, but no direct empirical testing has to my knowledge been done.

A number of more complex theoretical models have been recently published in order to account for the triggering of segmentation. The Inference-Based Event Segmentation account, for example, links segmentation with the process of inferring causality, i.e. the latent variable generating the observed situation (Shin & DuBrow, 2021), and the Information Optimisation Account homes in on the fact that troughs in predictability are themselves often predictable (Baldwin & Kosie, 2021). However, neither of these theories offer much in terms of direct empirical evidence, or in fact explain how the brain works out when the current event model should be abandoned. Also other, more computationally-oriented theories exist (e.g. Franklin et al., 2020; Khemlani et al., 2015), but their examination is outside the scope of this thesis, and not directly relevant to the present study.

In summary, it is still unclear what exactly triggers segmentation, as direct evidence of the necessity or sufficiency of prediction error is still missing and competing accounts do not offer testable hypotheses.

One way to begin to chart the exact mechanisms of segmentation is to turn to the brain responses that have been reported to occur at event boundaries. While there has been research done in the area for 20 years, there are not many studies that utilised methods similar enough to allow synthesis of results. In the next section I will review all the studies done between 2001 and 2019 that have reported localised neural activations in response to more or less naturalistic event boundaries.

1.1.3 Review of functional imaging studies

In this mini-review I will concentrate on studies that analysed brain responses to event boundaries in a continuous and extended stimulus. Only studies that utilised a behavioural segmentation paradigm or narrative shifts were included – studies where the events were defined in the experimental design itself, e.g. as separate short (< 20 second) film clips, were excluded. Also, as the interest of the present study is in boundary-related processing, all studies that did not report time-locked activations to event boundaries were omitted. I only review fMRI literature, as localisation of activations is rather unreliable in M/EEG and comparing those with fMRI results would be quite difficult. In the end, ten papers met the criteria for this review.

On first impression the literature presents a confusing abundance of boundary-related BOLD responses all over the cortex and in several subcortical areas. Yet, there is no single area which would be consistently activated in all of the studies – precuneus does appear in all results, but the shape of the response is sometimes unknown (Ezzyat & Davachi, 2011; Zacks et al., 2001, 2006), and at least once clearly not an activation peak (Speer et al., 2003). To make sense of this medley we need to take a closer look at several methodological choices made in the studies.

The studies can be grouped along several dimensions. Some dimensions pertain to the nature of the stimulus, which spans from extremely simple animations of abstract shapes (Zacks et al., 2006) to a full-length feature film (Ben-Yakov & Henson, 2018). Also modalities differ: many studies have used moving visuals with or without sound (Ben-Yakov & Henson, 2018; Speer et al., 2003; Swallow et al., 2011; Zacks et al., 2001, 2006, 2010), whereas others have utilised narrated stories presented via text (Ezzyat & Davachi, 2011; Speer et al., 2007, 2009) or audio (Whitney et al., 2009). The durations of the stimuli range from a few minutes (e.g. Zacks et al., 2001) to several hours (Ben-Yakov & Henson, 2018).

Another relevant stimulus dimension is the level of narrativeness. Early studies mainly used unedited, fixed shots of actors performing an everyday activity, e.g. doing the dishes, or even animations with abstract shapes, and these I will call non-narrative (Speer et al., 2003; Swallow et al., 2011; Zacks et al., 2001, 2006). Another few studies have used a semi-narrative account of everyday happenings, which feature an event structure but no dramatic plot (Ezzyat & Davachi, 2011; Speer et al., 2007, 2009). Finally, a handful of studies have utilised professionally produced narratives, which contain both event and dramatic structure and have not been significantly modified for research purposes (Ben-Yakov & Henson, 2018; Whitney et al., 2009; Zacks et al., 2010).

Another group of dimensions pertain to the segmentation method. Most studies have used some form of behavioural segmentation, with a group of participants annotating the stimulus more-or-less intuitively. On the extreme end is the study by Ben-Yakov and Henson (2018), where the annotators were given free reign with only a very general instruction ("where one event seems to end and another begin"). In contrast, many earlier studies instructed the annotators to segment on a specific grain, going as far as to train them with a test stimulus and giving feedback on deviating segmentation frequency (e.g. Zacks et al., 2006). It is also worth noting that with the exception of the study by Ben-Yakov and Henson (2018), no study reported having accounted for a reaction time (from boundary detection to button press) before analysing the behavioural results, though in most cases the annotators were not naïve to the stimulus and might have been able to anticipate the upcoming boundaries (for an examination of familiarity effects, see e.g. Michelmann et al., 2020).

A whole another group of studies have forgone the intuitive method altogether and instead coded the stimulus with narrative shifts, i.e. changes in the dimensions of the situational model (Ezzyat & Davachi, 2011; Speer et al., 2009; Whitney et al., 2009). All of these studies used verbal stimuli (either visual or spoken text) and defined boundaries as the beginning of the sentence which contains shifts in one or more narrative dimension. Even though there is a clear difference between this and the intuitive method, it is worth noting that intuitive boundaries often correlate with narrative shifts, and the more strongly the more dimensions shift at the same time (Speer et al., 2007; Zacks et al., 2010).

Finally, the stimulus, its duration and the instructions given to annotators together determine the average duration of the events used in the data analysis. These averages differ substantially between studies and might represent entirely different timescales of both neural processing and subjective experience. The shortest event durations are found in the studies using unedited movies of everyday activities, with fine-grain segments having mean duration of around 10 seconds, and coarse-grain about 30 seconds (e.g. Zacks et al., 2001). The longest durations are reported in studies using full narrative stimuli and especially when delivered via audio only: mean durations for coarse-grain segments may then exceed 1 minute (e.g. Whitney et al., 2009; Zacks et al., 2010).

In addition to these variables there are differences in whether the boundaries used for data analysis were gleaned from the imaging subjects themselves or from a separate annotator group, and whether the annotations were combined to yield an averaged set or not. However, these methodological choices seem to not affect the results in any meaningful way.

A final, highly meaningful difference between studies is their choice of analysis method. While more recent studies typically utilise the standard GLM analysis based on convolving the boundary timeline with the canonical HRF (see e.g. Huettel et al., 2014), the earlier studies instead collected averaged epochs around the boundaries and submitted the values at each timepoint to analysis-of-variance (ANOVA). This approach confounds HRF-type activation peaks with all other types of possible time-locked responses (e.g. activation troughs), and the results are thus not easily interpretable and do not allow direct comparison to results from GLM analyses. In some studies the shape of the response is plotted, and its relation to the canonical HRF can be estimated, but in some cases the authors have only reported the summary ANOVA results. These results I have generally treated with caution but have nevertheless considered in the following overview.

When these methodological dimensions are considered, some patterns of results begin to emerge. The clearest finding is that when the stimulus used exhibits narrative features (an event structure, at minimum), parts of the PMC – mainly PCC and precuneus – begin to respond consistently and robustly to event boundaries. The effect is modality-independent, and is found whether the stimulus is visual text (Ezzyat & Davachi, 2011; Speer et al., 2007, 2009), a film (Ben-Yakov & Henson, 2018; Zacks et al., 2010), or an audiobook (Whitney et al., 2009). The areas respond to boundaries defined by either intuitive segmentation or narrative shifts, and in the latter case they respond more strongly the more dimensions in the situation are changing (Speer et al., 2009). Further, activations in these areas are implicated in memory performance across event boundaries (Swallow et al., 2011). These responses are absent, however, in studies utilising simpler, non-narrative stimuli.

An opposite pattern is evident with activations in the superior temporal sulcus, especially in the posterior part (pSTS) in the right hemisphere: studies using movies with either human or abstract motion report it (Speer et al., 2003; Zacks et al., 2001, 2006, 2010), whereas studies using written or spoken narratives generally do not. Also, Speer and others (2009) found that areas in the STS responded specifically to shifts in characters and their goals, but not other situational dimensions, and the response did not linearly increase with the number of simultaneous narrative shifts, unlike in the PMC. It is thus unlikely that the activation peaks in pSTS are a general, modality-free marker of event segmentation. The same applies to other movement-related areas (e.g. MT+) which were a target of active and localised study in the early papers (Speer et al., 2003; Zacks et al., 2006), but have since received little confirmation in studies with more complex stimuli (see e.g. Zacks, 2020)

Another interesting finding that was already referred to above is the absence of consistent ACC activity. ACC modulations in response to boundaries are reported in three studies, and in only one of them the result definitely represents a transient activation (Speer et al., 2009). In the other two studies the response shape is either unknown (Ezzyat & Davachi, 2011) or resembles a transient deactivation (Zacks et al., 2010). The studies by Speer and others and Ezzayat and Davachi utilised semi-narrative texts, which were coded with narrative shifts, while the study by Zacks and colleagues featured an intuitively segmented film. Also, the study by Speer and others investigated specifically which brain areas exhibited activations that were linearly correlated with the number of concurrent situational changes, and thus did not in fact show that ACC in general would respond to boundaries – for a traditional segmentation-based analysis of the same data no response in ACC was found (Speer et al., 2007).

As event segmentation is known to affect and organise memory, the activity of the hippocampus and adjacent structures (mostly parahippocampal gyrus, PHCG) at event boundaries is of special theoretical interest. In the reviewed literature MTL responses are reported almost exclusively when hippocampus or PHCG are defined a priori as regions-of-interest (Ben-Yakov & Henson, 2018; Swallow et al., 2011) – the only exception being the study by Speer and others (2009), which found HC to be among the areas in which the response increased linearly as a function of the number of concurrent narrative shifts. Ben-Yakov and Henson found HC to activate in response to event boundaries in film stimuli, and Swallow and others found HC and PHCG to activate more strongly when subjects were attempting to recognise visual objects presented in a previous event as opposed to the current event. The absence of HC/PHCG responses in the other studies does not, however, necessarily mean that there would be no activation – especially the hippocampus proper is a challenging area for brain imaging methods, and its activity often remains unseen in whole-brain analyses (see next section for discussion).

In addition to these areas also the superior and middle temporal gyri, middle and inferior frontal gyri, temporo-parietal junction (TPJ), fusiform gyrus, and the lingual gyrus were reported as exhibiting boundary-related activity in more than one study. The superior temporal gyrus (STG) is exclusively present in studies using written text as stimuli (Ezzyat & Davachi, 2011; Speer et al., 2007, 2009), and TPJ and fusiform activations were only reported in studies using movie stimuli (Speer et al., 2003; Zacks et al., 2001, 2006, 2010). The middle temporal activations did not show a clear pattern, and neither did the frontal or lingual activations.

In summary, the most robust finding in the reviewed literature is the activation of parietal PMC areas (PCC and precuneus) in response to boundaries in narrative or semi-narrative stimuli regardless of stimulus modality. The pSTS, MT+, TPJ, and the fusiform gyrus were found to react to boundaries exclusively in movie stimuli, and STG activations were reported in studies using visual text. The hippocampus is found to respond when it is specifically targeted, but those studies within this research paradigm are very scarce. It is worth noting, though, that the PMC areas found to respond to event boundaries are also strongly linked to HC via anatomical and functional connections (Ritchey et al., 2015).

In the next section I will give a brief overview of what is known about the hippocampus and its role in declarative memory, and especially one-shot learning of events, also known as episodic memory. Finally, I will connect that with event segmentation and the emerging view of naturalistic memory processes.

1.2 Hippocampus and events

Event understanding is closely linked with episodic memory. Episodic memory, as traditionally defined, is memory of events in time and space, detailing the “what, where, and when” of experiences, and it gives rise to the subjective experience of remembering, thus crafting a link between memory and consciousness (Tulving, 1972). The crucial ability that enables all this is the capacity for one-shot-learning: a specific episode only takes place once, and it needs to be encoded there and then as an integrated whole to facilitate subsequent recall.

Since the early studies with patients suffering from medial temporal lobe (MTL) lesions it has been known that the hippocampus is essential for episodic memory encoding and explicit learning in general (Rempel-Clower et al., 1996; Squire et al., 2004; Squire & Zola-Morgan, 1988; Zola-Morgan et al., 1986). In this section I will review some of the theories and research that pertain to the role of hippocampus in memory, as well as outline some of the unique features of the hippocampus that make it both an especially important and a notoriously difficult brain area to study. Finally, I will review findings that link hippocampus to event segmentation.

1.2.1 Hippocampus as a memory encoder

The classical symptom of MTL damage is the “profound forgetfulness” of anterograde amnesia. A patient may be capable of functioning relatively normally in an ongoing situation, for example upholding an intelligent conversation, as long as there are no disruptions. But when the situation ends, no persistent memory trace gets formed, and the episode is promptly forgotten. The exact degree of the deficit is naturally dependent on the extent and location of damage sustained, but case studies going as far back as 1900 (reviewed in e.g. Squire & Wixted, 2011; Zola-Morgan et al., 1986), and studies on animal models of amnesia (reviewed in Squire et al., 2004) converge to support the clinical case for the necessity of MTL structures for declarative memory formation. Yet, it is less clear whether or not they are required for episodic retrieval. The pattern of retrograde amnesia in MTL damage suggests that over time episodic memories become independent of the HC – a process known as systems consolidation (Wixted & Cai, 2013) – but there have also been critical views (Winocur & Moscovitch, 2011).

In brain imaging studies on healthy human populations MTL in general and HC in particular have been found to be active during successful memory operations, both at encoding and retrieval, and over a variety of stimuli (Moscovitch et al., 2016; Paller & Wagner, 2002; Spaniol et al., 2009). Yet, it should be noted that HC is also implicated in general relational binding (Cohen et al., 1999; Olsen et al., 2012), and it is tricky to separate this function from memory encoding – in fact, complex visual scenes seem to elicit hippocampal activity more robustly than other types of stimuli (Henson, 2005). In other words, activations seen in imaging studies might at least partly stem from the processing of the stimulus and not purely episodic encoding.

Encoding an episodic memory requires the representation of the episode to be encoded as well as the ability to commit it to memory. The Memory as Reinstatement (MAR) model proposes that an experience is represented in the neocortex as a distributed "cortical neural pattern" (CNP) and simultaneously in MTL as a "hippocampal neural pattern" (HNP), which can be encoded and later reinstated, acting as a key to unlock the CNP and the full memory (Davachi & Danker, 2013).

The MAR model is built on evidence from functional neuroimaging studies, which have found reinstatement of cortical activation patterns during memory retrieval (Danker & Anderson, 2010), the degree of reinstatement correlating with the amount of recalled detail as well as subjective measures of vividness (C. M. Bird et al., 2015; St-Laurent et al., 2015). Also, as mentioned above, many recent studies utilising naturalistic stimuli have found event-bounded cortical activation patterns that are reinstated when the event is recalled and are even carried over to other brains through verbal communication (Baldassano et al., 2017; Chen et al., 2017; Zadbood et al., 2017). Thus, the existence of a situation-specific CNP is a well-documented phenomenon.

The case of HNP is also supported, though somewhat less clearly. Hippocampal activation patterns have been found to track situational features (Cohn-Sheehy et al., 2020; Milivojevic et al., 2016), and the degree of their reinstatement during recall is correlated with the level of detail retrieved, at least for coherent narratives (Cohn-Sheehy et al., 2020; Reagh & Ranganath, 2021). There is also evidence for hippocampal pattern completion and its role in the holistic reinstatement of the CNP during retrieval (Horner et al., 2015; Michelmann et al., 2020; Ritchey et al., 2013; B. P. Staresina et al., 2013).

But it might be misleading to conceptualise the hippocampal activations as merely a key to the cortical representation: the HC is implied in learning temporal sequences, which requires a representation that can distinguish time from content and thus differentiate between two events which are conceptually similar yet temporally distinct (Ranganath & Hsieh, 2016). The hippocampal code seems to also differentiate events that are close in time but not causally related, while connecting even distant events that together form a coherent whole (Cohn-Sheehy et al., 2020; Milivojevic et al., 2016; Bernhard P. Staresina & Davachi, 2009). Thus, the hippocampus seems to be building both walls and bridges with a complex logic that goes far beyond the tenets of MAR.

Further, it is debatable whether cortical pattern reactivation is indeed what makes a memory vivid: the transformation hypothesis suggests that it is rather the HC that is responsible for the sharp details and context of a recollection, whereas a memory depending solely on the neocortex is gist-like and semantic (Winocur & Moscovitch, 2011; see also Moscovitch et al., 2016). In other words, when a memory undergoes systems consolidation and becomes more independent of the HC, it necessarily loses some of its detail and specificity and transforms to something more schematised. There is also a considerable amount of research investigating the differential roles of the hippocampal subfields and the surrounding MTL structures in memory operations (see e.g. Henson, 2005; Ritchey et al., 2015), but that discussion is beyond the scope of this work.

In summary, it has been established that the hippocampus is necessary for memory encoding, and plays at least a large role in memory retrieval. So how then can the HC do what it does, and what the rest of the brain cannot, namely encode an integrated memory trace after just one exposure? In the next subsection I will briefly introduce some key findings from studies on both animals and humans, which have helped formulate our understanding of the mechanisms of hippocampal learning.

1.2.2 Plasticity and oscillations in hippocampal learning

Hippocampus is a notoriously difficult target for human functional brain imaging. For encephalography methods (M/EEG) it presents a challenge by being a deep structure, which leads to heavily attenuated signals (see e.g. Hari & Puce, 2017). Intracranial studies have been conducted (iEEG), but the availability of suitable surgical patients is a limiting factor. For magnetic resonance (MR) methods its deepness is also a problem, albeit a slightly smaller one, but as it is also a thin and long structure, with highly individually variable shape and orientation, it does not fare well in standard fMRI group-level analyses (Squire et al., 2004). Further, its proximity to sinuses often leads to signal drop-out in MR imaging, making particularly the anterior HC difficult to study (Asano et al., 2004; Greicius et al., 2003).

Thus, animal studies are an important source of insight into the operation of the hippocampus, as they allow not only its precise recording but also invasive manipulation of its function. Much of what we know of the HC is based on research conducted on rodents, which presents the problem of how those results transfer to the human hippocampus. Efforts are continually made to bridge that gap, but much remains unclear. Here I will present some key findings from animal research and how the phenomena have been explored in humans.

One key finding that originated from animal studies is that the HC is one of the very few brain areas that are capable of ongoing neurogeneration, and thus it retains a remarkable level of structural plasticity all through adulthood (Leuner & Gould, 2010). In other words, in the HC there is constantly a population of neurons in different stages of maturity, which possibly underlies its capacity for one-shot learning. It is also known to exhibit an exceptionally high density of NMDA receptors, which leads to easily triggered long-term potentiation (LTP) – animal studies have found that blocking LTP in the hippocampus with NMDA antagonists causes temporary anterograde amnesia, establishing its causal role (reviewed in Wixted & Cai, 2013).

Another key finding from animal studies are the hippocampal sharp-wave ripples (SWRs), which are tightly linked with memory and learning. They are bursts of high-frequency neural activity that take place during periods of awake immobility as well as slow-wave sleep, and they tend to reproduce the firing patterns elicited by a preceding experience in a speeded-up manner (reviewed in Bilkey & Jensen, 2021; Buzsáki, 2015). Especially post-experience SWRs replayed in forward order seem to be linked with memory consolidation, as disrupting them also disrupts learning (Ego-Stengel & Wilson, 2009; Girardeau et al., 2009; Jadhav et al., 2012) – SWRs are also detected prior to activity, which likely supports action planning (Pfeiffer & Foster, 2013; Singer et al., 2013), and post-learning in reverse order, which might relate to reward-related credit assignment (e.g. Ambrose et al., 2016).

Recently, SWRs and their concomitant rapid neural replay periods have also been studied in humans. For example, Liu and others used MEG to register the brain activity of participants during and in-between visual sequential learning tasks. They found that during the rest periods the data included bursts of spontaneous replay of the learned sequences, and the onset of the replay was marked by a high-frequency SWR burst (Liu et al., 2019). While source localisation in MEG is never perfectly reliable, the authors suggested that the replay component had neocortical sources, while the SWR originated from somewhere within the MTL. Another study by Sols and colleagues used EEG to detect similar rapid replay sequences following context shifts in image sequences (Sols et al., 2017), while Silva and others found the same effect in response to event boundaries in an audio-visual narrative (Silva et al., 2019), but neither study attempted to localise the source of the signal.

In summary, the hippocampus is unique in its propensity for plasticity, which likely underlies its capacity for one-shot learning. While "idling", i.e., between bouts of activity or during sleep, it is known to generate spontaneous rapid replays (SWRs) of preceding experiences, which is believed to play a role in consolidation of memories into LTM.

1.2.3 Hippocampal encoding at event boundaries

A traditional way of studying human memory encoding in the laboratory is to show volunteers a sequence of stimuli and later ask them to recognise them from another set of stimuli that consist of both old and new items (see e.g. Henson, 2005 for an overview of methodologies). Sometimes there may be additional questions to probe "source memory" and differentiate between vague familiarity and a more exact, episodic recollection. When studying memory encoding the subject is typically given a sham task and not told about the memory test, which makes it possible to study incidental encoding and gain some ecological validity for the results. Brain activations at the encoding stage are analysed based on the memory task performance to reveal subsequent memory effects: what differentiates a successful item encoding from a less successful one? This is often called the difference-in-memory paradigm (DM) (Paller & Mayes, 1987; see also Davachi & Danker, 2013).

Generally, the stimuli used in the traditional laboratory-based experiments of memory encoding in the hippocampus have been rather far from naturalistic episodic experiences. For example, Ranganath and others (Ranganath et al., 2004) presented their subjects with single words printed in either red or green, and the subsequent memory task required them to distinguish old from new items as well as recalling the colour of the word when it was presented (an indication of source memory). Other studies have used, e.g., face-name pairs (Kirwan & Stark, 2004), still pictures of everyday objects (Reas & Brewer, 2013), and still images of outdoor scenes (Stark & Okado, 2003). Source memory might be measured very indirectly, as in a pair rearrangement evaluation (Kirwan & Stark, 2004), or not at all (Stark & Okado, 2003).

The duration of the presented stimuli is typically brief. For example, the studies mentioned above had inter-trial intervals ranging (on average) from 0.5 to 7 seconds, and stimulus durations from 750 ms to 2.5 seconds. While these are entirely appropriate timings for studying perceptual processing, it is not altogether clear that the brain would really experience a two-second presentation of a single image as an *event* in a sense that it would trigger episodic encoding for every item separately, leading to measurable and memory-predicting activations.

These were some of the issues that informed a 2011 study by Ben-Yakov and Dudai, where the authors investigated the brain responses to short film clips in a DM paradigm (Ben-Yakov & Dudai, 2011). The film clips they used were disconnected and relatively short (4–16 sec), but still dynamic enough to share a resemblance to real-life episodes. Importantly, they paid close attention to the *eventfulness* of the clips, controlling the target clips with both scrambled and merely uneventful clips, which further focused the analysis on processing specific to event handling. Also, their memory task probed for the *gist* of the clip and required a full subjective recall of the stimulus, not merely an evaluation of familiarity.

The truly novel approach of the study was the hypothesis the authors had concerning the timing of the response. They realised that if we assume that the content of the memory engram is a relationally and sequentially organised whole, we should expect that a critical part of the encoding would take place only *after* the event has ended, when all the information comprising the episode is available in the working event model. The authors analysed separately the responses occurring during the presentation of the clip – the intra-stimulus or “online” encoding – from what happened right after the clip ended – the delayed or “offline” encoding.

What they found was that activity in distinctly different brain areas predicted offline and online DM. Especially the bilateral hippocampus and dorsal striatum (caudate nucleus) seemed to be specifically activated by clip offsets, and their response was significantly larger for remembered than forgotten clips. But critically, this activity took place *only* after eventful clips: uneventful and scrambled clips did not elicit a similar response. The authors concluded that this effect, especially in the hippocampus, reflects the binding and consolidation of an integrated episodic representation.

Later the authors showed that this hippocampal response is indeed related to the offset of the clip, and not just the onset of the rest phase, as it takes place even if the first clip is immediately followed by another (Ben-Yakov et al., 2013). The response to the first clip was diminished, which was shown to be an indicator of retroactive interference, as the amount of decrease was correlated with poorer subsequent memory. It is likely that the design in this study did not give the participants sufficient forewarning of the impending end of the clip to allow for adequate consolidation – in longer, naturalistic narratives creators take care to craft event boundaries that take this into account. Anyhow, the authors proposed that this hippocampal activity peak is driven by event boundaries during naturalistic episodic encoding and could possibly reflect a rapid replay of the preceding episode, which has been found to take place in the rodent hippocampus.

They also showed the existence of an onset-related hippocampal activation (especially in the right posterior hippocampus), which was not triggered by novel events, but emerged as the stimulus was presented repeatedly during the experiment and became familiar (Ben-Yakov et al., 2014). Concurrently the hippocampal offset response was attenuated.

Because of these findings the study by Baldassano and others in 2017, which examined the neurally defined cortical events, also searched for hippocampal activation peaks time-locked to cortical pattern shifts. They found that in the AG, PMC and PHCG pattern shifts tended to also trigger a hippocampal response, which the authors interpreted as being an indication of memory encoding (Baldassano et al., 2017). Based on their results they suggested a model in which the cortical areas build and maintain the online representations, with the highest regions implementing what we consider the working event model, and where the occurrence of event boundaries triggers the hippocampus to encode the event into memory, which is seen as the post-event activation peak.

These findings, together with what was known of the involvement of the hippocampus in general relational binding (Olsen et al., 2012), led Richmond and Zacks (2017) to formulate a dual-role model for hippocampal involvement in event segmentation and encoding. They proposed that the hippocampus is continuously binding elements of the ongoing situation into a working model, but the offset of an event triggers an additional burst of activity which serves to sharpen the bound pattern in a way that enables its subsequent retrieval, even after it has been replaced by a new situation and a new binding in the hippocampal workspace (Figure 2). Thus, a weaker activity peak might lead to less robust encoding, and thus less successful recall, which is in line with the DM results reported by Ben-Yakov and others. The increased hippocampal activity at the onset of familiar events, and generally when retrieving episodes from long-term memory (McDermott et al., 2009; Spaniol et al., 2009), would in turn be related to the increased material available for the general binding function.

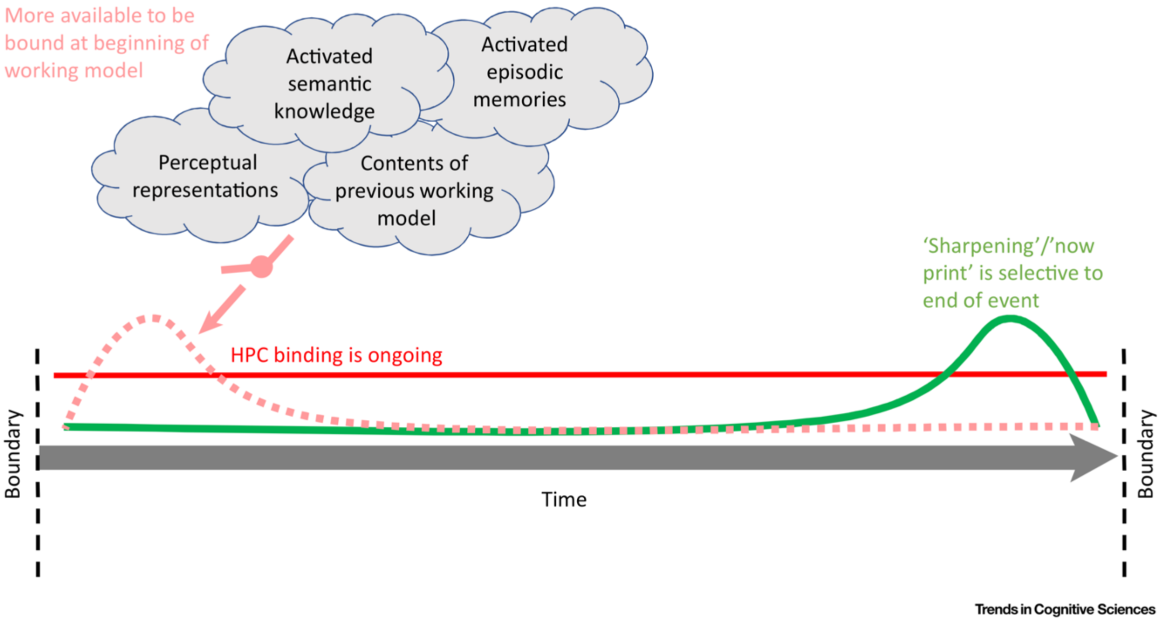


Figure 2: The dual-role model for hippocampal involvement in event memory. Relational binding in the hippocampus is constant and obligatory, but more material available at the beginning of a new event translates into an activation peak. At the end of the event the bound representation is "sharpened" with increased activation to facilitate its encoding in to LTM (Richmond & Zacks, 2017).

Up until 2018, however, there was no clear empirical evidence that event boundaries in a naturalistic experience actually do trigger hippocampal activity peaks. This had not been directly tested by Baldassano and colleagues (2017), and neither had hippocampal peaks been reported in any studies on event boundaries – the latter possibly due to the difficulty of detecting hippocampal signals with fMRI. In 2018 Ben-Yakov and Henson filled the gap by studying the hippocampal response to behaviourally defined event boundaries in two film stimuli (Alfred Hitchcock's "Bang! You're Dead", edited into an 8 minute version, and Robert Zemeckis's Forrest Gump, a slightly edited 2 hour version) (Ben-Yakov & Henson, 2018). The behavioural segmentation was done by a separate group of annotators, who got the simple and intuitive instruction of marking the places where they felt that "one event ended and another began", with no training or guidance with regard to segmentation grain, and no second pass. This allowed the authors to identify boundaries that were minimally affected by the assumptions of the researchers, and get as close as possible to the natural, subjective experience of events as they are encountered. The analysis of the imaging data concerning the event boundaries was restricted to the hippocampus, with its location manually traced from anatomical images.

The results showed a significant hippocampal response to boundaries in both datasets. The authors also extracted hippocampal peaks in a data-driven manner and showed that they significantly lined up with the behavioural boundaries – though it should be noted that around half of the peaks in both datasets (42% and 62%) did not match with a boundary, and it remained unclear what the driving feature for them might be.

In order to investigate what features of the stimulus might be relevant for the hippocampal activity – i.e., what does an “event” or “boundary” for hippocampus in a naturalistic stimulus actually consist of – the authors also quantified a large number of visual and auditory features, as well as spatial and temporal narrative shifts and responses from primary sensory areas, and tested them together with the behavioural “salience” of the boundaries (estimated from the number or annotators marking a specific boundary) to assess their relative contributions. Running these models in the hippocampus as well as in the cortical ROIs of the Harvard-Oxford atlas they found that only in the hippocampus and the posterior cingulate the modulation of the response magnitude by behavioural salience was not fully explained by the lower-level features of the stimulus. While this does not mean that the hippocampal activation would be independent of low-level stimulus features, it nevertheless implies that the hippocampus might be working on representations that are very high up in the abstraction hierarchy of the brain, and that the effects are not merely a consequence of the known hippocampal sensitivity to time and space.

In summary, hippocampus is known to be an integral player in episodic encoding and retrieval, and more recently it has been shown to activate specifically in response to event offsets. It has been suggested that this activation peak would represent a sharpening of the event pattern, and possibly consist of a rapid neural replay of the experience (Ben-Yakov et al., 2013; Bilkey & Jensen, 2021). It has been suggested that hippocampal activation peaks might be the domain-general neural marker of event segmentation in a naturalistic experience (Richmond & Zacks, 2017). However, empirical evidence for this is still rather scarce, and limited to stimuli in the audio-visual modality.

1.3 Aims of the study

To summarise, event segmentation is a demonstrably automatic process, which chunks our ongoing experience and affects how it is perceived, encoded, and retrieved. The model for the current event is likely implemented as a distributed representation across many high-level cortical regions, most of which belong to the DMN. Further, event boundaries seem to have a unique role in event processing and memory, as they are likely to be the points at which a working event model is transformed into a persistent memory trace, and then discarded. This feat of episodic encoding is very likely to involve the hippocampus, and evidence of its transient activation at event boundaries does exist (Ben-Yakov & Henson, 2018).

However, these activations were found in response to two films, i.e. audio-visual stimuli. This still leaves open the possibility that the activations are related to visuospatial features of the stimulus, as the hippocampus is known to be involved also with ongoing visual binding and spatial cognition. Ben-Yakov and Henson (2018) did investigate the effects of visual, auditory and spatial features to the hippocampal response magnitude and concluded that those did not explain all of the response, but another way of ascertaining the independence from visual features is to use a stimulus that does not have a visual component at all.

Thus, for this study, I use data gathered from participants passively listening to an extended auditory narrative. If the hippocampal boundary response is truly a domain-general marker of segmentation in naturalistic experience, then we would expect to find it in response to event boundaries in our audiobook. Also, it should not be driven by structural components smaller than an event, i.e., the response should be significantly higher to event boundaries than to non-event sentence boundaries.

In addition to the main question of hippocampal boundary responses I will do an exploratory analysis over all cortical regions-of-interest (ROI). The variety of cortical areas that have been linked to boundary processing is extensive, covering much of the associative cortex (i.e., brain regions not occupied with low-level sensory processing). However, the majority of studies investigating boundary-related responses in the brain have been conducted using visual stimuli: either text, dynamic visual presentations or audio-visual narratives. Only one previous study utilised an auditory narrative, and there the authors found a rather more restricted set of cortical activations – namely areas in precuneus, PCC and middle cingulate (Whitney et al., 2009). This raises the possibility that the responses in precuneus and PCC might be the most domain-general ones, while the lateral cortical activations might have more to do with visual features of the stimulus. However, the study by Whitney and colleagues did not use an intuitive segmentation paradigm, but instead coded the stimulus with narrative shifts, which might affect the results.

I will also investigate the effects of both behavioural salience and lower-level auditory features of the stimulus on the response magnitude in both hippocampus and the cortical ROIs. This is done in order to differentiate between areas that are genuinely sensitive to boundary salience from those that are in fact more attuned to speech onsets, offsets, or pauses, and it will give a complementary view on the boundary-sensitive areas. Following Ben-Yakov and Henson (2018) we would except to see significant salience modulation in the hippocampus and the PCC.

Further, as it known that familiarity with a given event may diminish the need for encoding and thus reduce the offset response in the anterior hippocampus (Ben-Yakov et al., 2014), we would expect this to happen with participants listening to our auditory story for the second time. Our data consists of fMRI recordings from both first and second time listeners and enables us to test the effects of familiarity.

For the segmentation method I will use a naïve and purely intuitive segmentation paradigm to 1) mirror the methods of Ben-Yakov and Henson (2018) as closely as possible and facilitate comparison, 2) to augment the findings of Whitney and others (2009) with a behaviourally segmented stimulus, and 3) to base our findings and discussions on an ecologically valid foundation of subjective experience.

Note that none of the segmentation studies reviewed above reported hippocampal activations, which is quite possibly due to the demanding nature of HC as an imaging target that puts it at a disadvantage in whole-brain (esp. univariate) analyses. Thus I will not attempt to analyse the hippocampal findings together with the cortical ROIs, but instead target it separately, as more refined methods of extracting and analysing hippocampal fMRI signals (see e.g. Squire et al., 2004) are beyond the scope of this thesis.

So, to summarise the hypotheses:

1. We expect to see boundary-related transient activations in the hippocampus, and we expect them to be larger for event boundaries than for non-event sentence boundaries, and to be modulated by boundary salience even after controlling for low-level audio features.
2. We expect to see cortical activations in PMC, namely the precuneus and the PCC, and we expect not to see lateral or frontal activations. We also expect that the response in the PCC will be modulated by boundary salience, similarly to the hippocampus.
3. We expect first time listeners to exhibit a stronger hippocampal response to event boundaries than second time listeners.

2 Materials and methods

2.1.1 Participants

This study analysed an fMRI dataset collected from participants listening passively to an audiobook (Hakonen et al., 2020). The data was recorded from 50 participants (25 females and 25 males, 18–35 years, mean age 24.6 years). All participants were fluent in Finnish (either native or otherwise highly fluent) and had given their informed consent before participation. The study was approved by the Aalto University Research Ethics Committee and conducted in accordance with the Declaration of Helsinki.

The behavioural boundary data was collected from a separate sample of 20 participants in two parts: first with in-lab experiments (N=8) using the PsychoPy software (Peirce et al., 2019), and later with an online experiment (N=12) using the Pavlovia platform (https://pavlovia.org). The move to the online platform was due to the insufficiency of the first sample and the pandemic situation caused by the covid-19 virus, which prohibited the use of in-lab experiments. In the online experiment 3 participants had to be excluded due to platform malfunction, which resulted in all data being lost. Thus the final behavioural data set included data from 17 participants (11 females, 4 males, 2 other). None of the participants reported a history of neurological or psychiatric disorders, and all were native Finnish speakers. All participants gave their informed consent before participation and were compensated for their time with one movie ticket.

2.1.2 Stimulus and segmentation procedure

The stimulus was a 71-minute-long audio narrative, written and narrated in Finnish. The story was originally devised for a study investigating effects of shared family cultural background on brain responses during narrative processing (Hakonen et al., 2020), but for the purposes of this study it can be considered naturalistic, as it incorporated no experimental manipulation of event boundaries and was generally produced to play like a naturalistic narrative. The story was presented to the subjects in 10 parts in both the behavioural and the brain-imaging experiments (4.78–8.42 min, mean length 7 min).

The participants in the brain-imaging experiment listened passively to the narrative while in the scanner. A white fixation cross was shown during the audio stimulus, and still images of Helsinki were displayed during reference scans and breaks. Each chapter was presented during a single fMRI run, with a possibility of breaks in-between. All participants heard the story once in the MRI scanner and once during MEG/EEG acquisition. The sessions were separated by at least one month, and their order was counterbalanced across genders. Thus, half of the participants were 1st time listeners (heard the story first in the MRI scanner) and the other half were 2nd time listeners (had already heard the story once before coming in for the fMRI session). The groups are henceforth termed LIS1 and LIS2, respectively, with the LIS2 group having relatively more familiarity with the story, which might affect both segmentation schedule and the hippocampal response size. The MEG/EEG data was not used in this study.

In the behavioural experiment the participants were instructed to segment the stimulus by pressing a key on the computer keyboard whenever they felt that one event had ended and another begun. No instruction was given regarding the grain of segmentation. Participants were given the definition of event as a “period of time with an end and a beginning”, and instructed to follow their intuitive feeling of discontinuity, as might be perceived e.g. between morning routines and being at work.

The participants had to make their decisions “online”, i.e. while listening to the story for the first time. The audio could not be paused or rewound, and the participants could take breaks during the experiment only at the predefined chapter boundaries.

Some of the in-lab participants (N=6) were asked to make a second run on their annotations and adjust them to reflect the points in time where they first detected the boundary to remove the reaction time from detection to pressing the button (adjustment task). These modified annotations were later used to estimate an average reaction time specific to this stimulus.

2.1.3 fMRI acquisition and preprocessing

The MR imaging was performed with a 3T MRI whole-body scanner (MAGNETOM Skyra, Siemens Healthcare, Erlangen, Germany) using a 32-channel receiving head coil array. Anatomical images were measured using a T_1-weighted MPRAGE sequence (TR=2530 ms, TE = 3.30 ms, field of view (FOV) = 256 mm, flip angle = 7 degrees, slice thickness = 1 mm). Whole-brain fMRI data was measured using an ultra-fast simultaneous multislice (SMS) inverse imaging (InI) sequence (Hsu et al., 2017), with a TR of 0.1 seconds. Other imaging parameters for the functional runs were TE = 27.5 ms, flip angle = 30°, FOV = 210 x 210 x 210 mm^3, slice thickness = 7 mm, and in-plane resolution = 5 mm x 5 mm. Solving the inverse problem in InI reconstruction requires a sensitivity map of the channels in the coil array (Hsu et al., 2017) – this information was included in a 6-sec reference scan measured before each functional run with the same imaging parameters as used in functional imaging. For full description of the imaging procedure and image reconstructions, see Hakonen et al. (2020).

Each functional run started with a 12.3-second period without auditory stimulation, which was removed from the data prior to preprocessing. After the end of the stimulus the registration was continued for another 14.7 seconds to allow for the BOLD response to level out, and this tail signal was retained in the final data.

In preprocessing the functional images were co-registered to the MNI152 standard space with 3-mm resolution, and the scanner drift was removed with a Savizky-Golay filter (order = 3, frame length = 240 s). Physiological and movement artefacts were suppressed with the MaxCorr method (Pamilo et al., 2015) by regressing out 10 components which correlated maximally within the white matter and cerebrospinal fluid of the participant, but minimally within the same areas of other participants. The data was then filtered between 0.08 and 4 Hz with a zero-phase filter, and finally spatially smoothed with a Gaussian kernel (FWHM = 6 mm). For the purposes of this study the temporal resolution of 100 ms was impractically high, so the data was further temporally smoothed with a window of 2 seconds and a 50% overlap to obtain a dataset with a TR of 1 second. Due to the windowing the temporally smoothed data began 1 second later than the original and ended 1 second earlier, resulting in a dataset where the runs were 2 seconds shorter than in the high-resolution data.

The region-of-interest (ROI) time courses were extracted and averaged using Python 3.9 and the scikit-learn/nilearn packages (Abraham et al., 2014), using the Harvard-Oxford cortical atlas (thresholded at 25%) and the hippocampal mask from the Harvard-Oxford subcortical atlas (thresholded at 25%). The time courses were detrended and z-scored, and the data from all 10 functional runs was concatenated. The total length of data per participant was thus 4335 TRs, i.e. 4335 seconds. In total, 49 ROIs were analysed.

All computational resources were provided by the Aalto Science-IT project.

2.1.4 Processing of boundary annotations

As the boundary annotations were done naïvely and intuitively it was likely that this resulted in some lag between boundary detection and button press. To account for this a reaction time was estimated based on the data from the six participants that had completed the adjustment task. Both boundary sets from each participant were examined side-by-side and unpaired boundaries removed. The difference between each original-adjusted pair was then calculated, and an average taken across all boundaries from all participants. The reaction time calculated with this method was 2.93 seconds, which was rounded up to 3 seconds, and this was subtracted from all annotations – if the subtraction resulted in a negative value, it was corrected to zero. In addition to this, another dataset with a 1 second reaction time was created and used for the analyses, to mirror the methods of Ben-Yakov & Henson (2018) – though again rounding up to a full second. Henceforth these boundary sets are termed the *latency -3s* and *latency -1s* sets.

After the removal of the reaction times the annotations made by different participants were combined following broadly the method of Ben-Yakov & Henson, 2018): In the first pass all individual annotations were chained into a cluster, as long as they were less than 1.5 seconds apart, and in the second pass these clusters were in turn chained into 2nd level clusters, as long as their mean time points were less than 4 seconds apart. In between the passes all clusters with just one participant were dropped, and also in the 2nd level clusters a single participant could only be counted in once. The final time point of a boundary was defined as the average of the individual annotations forming the 2nd level cluster, with nObs defining the number of participants that marked it. Finally, all 2nd level clusters with nObs less than 5 were dropped.

The time windows and thresholds were selected by trial and error so as to obtain a final boundary set which generally matched the descriptive statistics across the participants, i.e., the mean number and duration of segments. Also the content of the stimulus was taken into account: it turned out that the event structure of the audio story used here was somewhat diffuse, and the boundaries sometimes evolved over the course of one or even several sentences, resulting in annotations sprinkled over the whole duration. Thus, the “cluster lengths” of some boundaries tended to grow long. Yet, they were genuine boundaries, and it seemed sensible not to let them splinter, even if that resulted in some inaccuracy in the boundary timings. For this reason rather long clusters were allowed in the combined set, though even the longest cluster was kept shorter than the mean minimum segment length from the original annotations (i.e., the mean of the shortest segments across the participants).

After clustering a further set of boundaries was added post hoc to the boundary sets, to denote the beginnings and ends of the chapters (speech onset and offset, respectively). This was deemed necessary, as the chapter borders also represented major event boundaries, but due to the design decision of having the annotators listen to the stimulus chapter-by-chapter they had not consistently marked those boundaries.

2.2 Data analysis

2.2.1 Boundary response in the hippocampus and cortical ROIs

The main interest of this study was the replication of the hippocampal boundary-related response, and for this I followed the permutation-based method used in several previous studies (e.g., Baldassano et al., 2017; Ben-Yakov & Henson, 2018).

This method has the advantage of not relying on assumptions concerning probability distributions as it models them directly from the data – albeit requiring more computational power than traditional analysis methods. The BOLD response to boundaries was measured by fitting a GLM with one predictor for the event boundaries (disregarding their nObs), and 13 nuisance predictors (three polynomial drift regressors defined per run, and constants indicating the runs). The event predictor was convolved with the canonical HRF. The design matrix was defined with the nilearn module, and the GLM was fitted to the data with the OLS function in the statsmodels module (Seabold & Perktold, 2010). The resulting betas were then averaged over participants.

To assess the significance of the response a sample of 1000 boundary permutations was created by randomly shuffling the event segments (i.e., the intervals between consecutive boundaries), and the proportion of permutations eliciting a higher average response than the intact set was used to derive a p-value (1-tailed, perms > bounds, betas for each permutation averaged over participants).

The same analysis steps were performed on the data from the hippocampal ROIs as well as the cortical ROIs. In the latter case the results were corrected for multiple comparisons with the Holm-Bonferroni method (Holm, 1979). The results from the hippocampus were not adjusted. The results were calculated for both boundary latencies (-3s and -1s).

2.2.2 Familiarity effects

Earlier studies have shown that the hippocampal response to boundaries may also be modulated by stimulus familiarity (Ben-Yakov et al., 2014). To assess the effect of familiarity in this data, the difference between the mean boundary response in the familiarity groups (LIS1/LIS2) was tested. Subject-wise beta values for the latency -3s boundaries were collected, and the group means were tested with a Welch two-sample t-test (2-tailed) and effect sizes estimated with Cohen’s d, with the R package effsize (Torchiano, 2020).

2.2.3 Controlling for audio gaps and sentence boundaries

As the stimulus used in the current study was an audiobook, it seemed prudent to also control for the underlying linguistic structure. For this purpose I adopted the idea of the method used by Whitney and others (2009).

First, all speech gaps longer than 1 second were collected and individually checked in terms of whether they were sentence boundaries or not. After dropping six non-sentence boundaries a pool of 300 speech gaps remained. These were then checked against the latency -3s event boundaries: if a gap was in the proximity of an event boundary (±2 sec from offset and onset of gap), it was marked as boundary-related, which resulted in 52 *event gaps*. One further gap was dropped due to being too close to an event gap. The remaining 247 gaps comprised the pool of *non-event gaps*, from which 1000 random samples of 52 gaps were collected. With this permutation method the segment lengths in the sham sets could not be strictly controlled for, but they were restricted to be more than 6 seconds, i.e., the sampling was redone until the whole sham set fulfilled the criterion. The resulting set was visually inspected to ascertain randomness. For all gaps the middle of each gap was used as the timing for the presumed neural event.

The event gaps and sham non-event sets were used to fit separate GLMs (with identical confounds, as before), and the proportion of sham beta-values higher than the one elicited by the event gaps was taken as the p-value (1-tailed test). The analyses were run on the hippocampus and in all cortical ROIs – across the cortical areas the p-values were again corrected with the Holm-Bonferroni method.

2.2.4 Salience modulation

Another central question of the present study was whether the boundary-related responses in the hippocampus and the cortical ROIs could be explained by the low-level features of the stimulus, or whether they were sensitive to perceived salience even after taking perceptual confounds into account. Salience was defined as the number of observers (*nObs*) who had marked the boundary, binned into three salience bins (low/med/high) of relatively equal sizes.

To account for the effect of perceptual features two confounds were calculated for each boundary in the latency -3s set: *meanVol* was defined as the mean volume (as percentage of maximum volume) in a 1 second epoch around on the boundary, and *volDiff* as the difference between the 500 ms halves, so that values above zero represented increasing volume. These were then entered into a mixed-effects model following the simplified method introduced by Westfall and others (Westfall et al., 2017): a GLM with one regressor for each event was fit to the fMRI data (a beta-series approach, Rissman et al., 2004), and the resulting beta values were fed into a mixed-effects model with salience and perceptual confounds as the fixed effects, and event and subject ids as random effects (*betas ~ salience + meanVol + volDiff + (1 | event id) + (1 | subj id)).* The model was fitted and tested with the lme4 and lmerTest packages in R (Bates et al., 2015; Kuznetsova et al., 2017).

For this analysis the post hoc boundaries at chapter borders were modelled in the GLM but not entered into the mixed-effects model. The model was run on the hippocampus and all cortical ROIs, and the p-values were corrected for multiple comparisons with the Holm-Bonferroni method to assess whether the significance of the salience modulation was as specific to the hippocampus – and possibly PCC – as it was in Ben-Yakov and Henson (2018).

2.2.5 Plotting

For plotting purposes finite impulse response (FIR) functions were used to model the shape and timing of the BOLD response (Ben-Yakov & Henson, 2018; Huettel et al., 2014, 384–385). Specifically, each time point in the range [-5,10] relative to the boundary time point was represented by a separate stick function, resulting in separate beta estimates. The betas were then averaged over participants and plotted to illustrate the mean response.

Because the boundaries themselves had to be shifted -5 seconds in order to plot negative delays, some boundaries crossed the beginning of their respective parts, and had to be dropped. For the statistical analysis – on familiarity or salience modulation – the whole set could be used.

All plots were produced with R-4.0.0 and RStudio 1.4 (R Core Team, 2020; RStudio Team, 2019) using the package ggplot2 (Wickham, 2016).

3 Results

3.1 Behavioural results

As the grain of segmentation was not predefined, the participants in the behavioural experiment came up with a highly variable numbers of annotations (24–222, mean=83.35, median=63). However, no participant was judged to be an outlier, as all were inside 2.41 standard deviations from the mean, so all data was used to calculate the combined boundary set. The bridge lengths and thresholds used in the procedure were selected so as to result in a total number of boundaries between the mean and median of that from individual participants, as well as similar mean segment lengths, while avoiding overly long clusters (i.e., longer than 10 seconds; see Table 1). To account for the behavioural lag from boundary detection to button press a reaction time was removed from all annotations: I used both a 1 second reaction time based on the methods of Ben-Yakov and Henson (2018), and also a 3 second reaction time specifically estimated for the current stimulus.

The combination procedure resulted in a set of 68 boundaries for the latency -1s condition and 66 for the latency -3s condition, as two boundaries in the earlier set had to be dropped due to the shortening of the neural data following from the temporal smoothing procedure. After clustering some post hoc boundaries were added to denote the beginnings and ends of the chapters, which resulted in a final set of 82 boundaries for the latency -1 condition, and 79 for the latency -3s condition (Figure 3).

Table 1: Statistics of the behavioural data before and after clustering. Segment lengths are the lengths of events between the boundaries, and cluster lengths are the span of time between the first and last original time point contained in the cluster. Segment lengths per participant are mean of means and median of medians.

	nBounds	*segment lengths*	*cluster lengths*
per participant	median = 63 mean = 83	median = 58.0 sec mean = 75.3 sec	
final boundaries	68	mean = 61.66 sec sd = 43.58 range = 7.22–206.02	mean = 4.34 sec sd = 1.89 range = 0.82–9.18

Table 2: The final clustered boundaries (without post hoc boundaries) for both latencies tabulated by number of observers marking each boundary.

nObs	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**total**
nBounds -1s	15	6	11	5	6	7	3	5	0	4	3	2	1	68
nBounds -3s	14	6	10	5	6	7	3	5	0	4	3	2	1	66

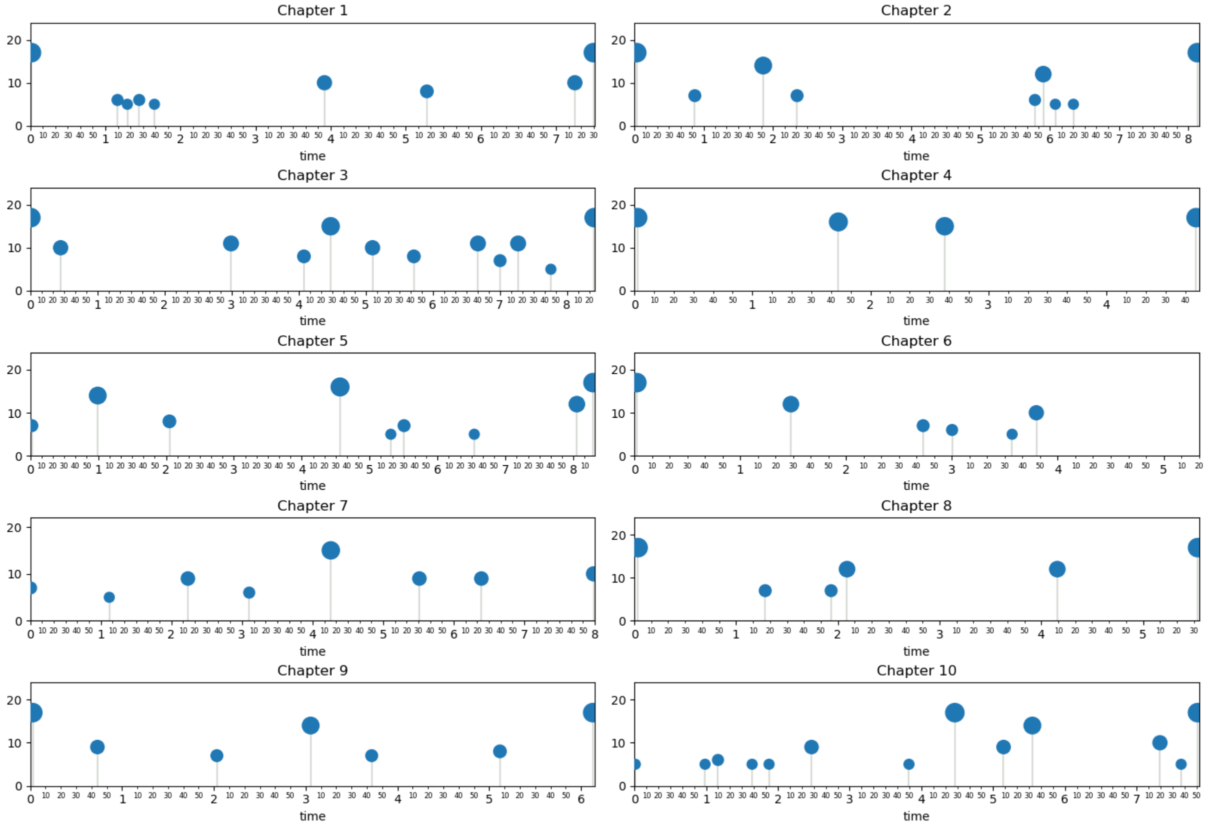


Figure 3: Final boundaries of the latency -3s condition, plotted on each separate chapter. Both height and size of balloon denotes nObs, i.e. the number of participants that were included in that cluster. The large nObs=17 boundaries at the beginnings and ends of some chapters were added post hoc to denote the chapter limits, which were not consistently marked by the annotators, but would still be assumed to elicit a boundary-related response.

3.2 Hippocampus responds to event boundaries

The main question of this study was whether or not the hippocampus responds to behaviourally defined event boundaries in an auditory narrative. This was analysed for two different boundary latencies: one with a 1 second reaction time removed, and the other with a 3 second reaction time. Significance was tested by comparing the beta value from the intact boundary set to beta values for a sample of 1000 scrambled sets, in which the segment lengths were kept unaltered, but their order was permuted. The portion of permuted sets with a beta higher than that of the intact set was taken as the p-value.

The hippocampal response was significant for the earlier latency, but not for the later one (latency -3: beta=13.267, p=0.002; latency -1: beta=5.500, p=0.115; see Figure 4). From the estimation of response shape (Figure 6) it seems clear that the earlier latency is a better fit for the time course of the activation, which might imply that the behavioural segmentation of this stimulus – and perhaps auditory narratives in general – is somewhat more difficult than that of visual or audio-visual stimuli, and results in longer lags.

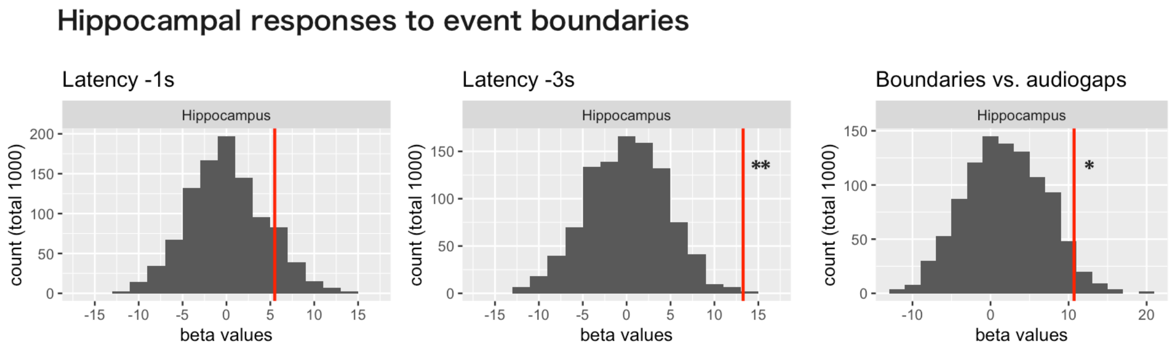


Figure 4: Hippocampal response to boundaries (red line) vs. permutations (histogram) with both latencies as well as controlled against speech gaps. Latencies are given relative to the behavioural boundary, with -1 based on previous studies and -3 estimated for this stimulus. Latency -1s beta=5.500 (p=0.115), latency -3s beta=13.267 (p=0.002), and boundaries vs. gaps beta=10.699 (p=0.038).

For the main analyses I wanted to follow the methods outlined by Ben-Yakov and Henson (2018), but as the auditory stimulus in this study differs in many respects from the Hollywood films they used, some additional controls were deemed necessary. For instance, it is a plausible argument that the activations seen here in response to event boundaries are, in fact, explained by their larger-than-chance coincidence with pauses in speech or the occurrence of a sentence boundary. Especially for the latency -3 condition this is a serious confound.

For this reason I did a further analysis by selecting all the speech gaps (with the minimum duration of 1 second) that were also sentence boundaries, which gave a pool of 300 lower-level boundaries, 52 of which were associated with an event boundary. The event-related speech gaps were then contrasted with random samples from the non-event speech gaps. Even with this control the hippocampal response to event boundaries remained significant (beta=10.699, p=0.038; Figure 4).

It should be noted, however, that in our stimulus the durations of the speech gaps varied naturalistically due to the rhythmic choices of the narrator, and thus could not be controlled for. In fact, the event-related gaps were on average somewhat longer than non-event gaps (event mean=1.858s, sd=0.714; non-event mean=1.332s, sd=0.253; t=5.239, p<0.001). In other words, listeners might have taken their cues for segmentation based on the speaker's rhythm, and not exclusively on abstract narrative features.

Earlier studies have shown that the hippocampal response to boundaries may also be modulated by stimulus familiarity (Ben-Yakov et al., 2014). To test this, I also compared the hippocampal response magnitude at latency -3 sec between the first and second time listeners of the story. No significant differences were found (t=0.291, p=0.77; see Figure 5).

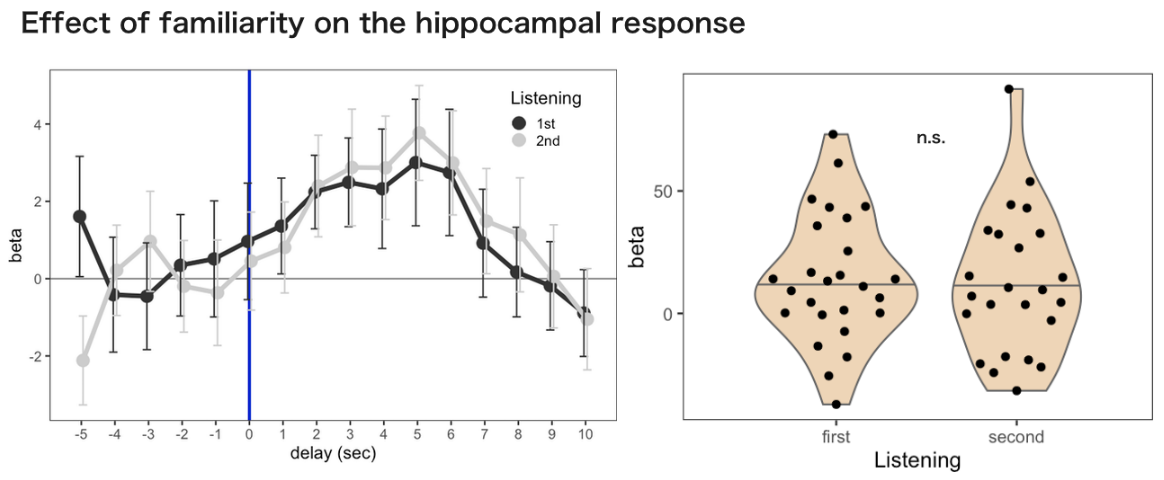


Figure 5: Effect of familiarity on the hippocampal response for the latency -3s boundaries. Difference between responses in the two groups (1st and 2nd time listeners) is not significant.

3.3 Event boundaries engage the PMC and ACC

The main analyses were also run on all the areas from the Harvard-Oxford cortical atlas, resulting in several areas that responded to event boundaries with both latencies, and some which only responded to the earlier latency (see Table 3 and Figure 6).

The most robust responses for both latencies were seen in the PCC, which is well established in the literature as being sensitive to narrative boundaries in various modalities. Another typical finding was the precuneus, which in this data responds more strongly to the later timing and can be seen (Figure 6) to have a slightly delayed response profile compared with the PCC. The posterior PHCG has a similar profile. Somewhat surprisingly there were also strong responses to both latencies in all medial visual areas, namely lingual gyrus, the calcarine cortices and the cuneus. This is especially curious as the design in the present study included no visual stimulation during the data runs.

Some of the cortical ROIs responded significantly only to the earlier boundaries (latency -3): in frontal operculum, insula and ACC the response to high salience boundaries peaked at 3 seconds after the behavioural timing. In addition to these there were also boundary related responses in the vmPFC (termed frontal medial cortex in the HO atlas. It is noteworthy that apart from insula no lateral ROIs were found to respond to boundaries with either latency, which stands in contrast to findings from earlier studies on event boundaries.

Table 3: Betas from ROI analyses on intact boundaries. The p-values were obtained with the same method as with the hippocampal response (portion of perms > bound) and corrected for multiple comparisons with the Holm-Bonferroni method. The table contains all areas with corrected p < 0.05. Areas and labels are from the Harvard-Oxford cortical atlas.

Latency -3 sec			**Latency -1 sec**		
area	beta	p (adj.)	area	beta	p (adj.)
Cingulate Gyrus, posterior division	33.165	***	*Cingulate Gyrus, posterior division*	32.058	***
Supracalcarine Cortex	31.986	***	*Precuneous Cortex*	25.722	***
Intracalcarine Cortex	29.305	***	*Supracalcarine Cortex*	25.574	***
Cuneal Cortex	27.721	***	*Intracalcarine Cortex*	21.73	***
*Cingulate Gyrus, anterior division **	26.188	***	*Cuneal Cortex*	21.219	***
Precuneous Cortex	24.281	***	*Parahippocampal gyrus, posterior*	19.071	***
Lingual Gyrus	24.261	***	*Lingual Gyrus*	16.974	***
*Insular Cortex **	21.53	***	*Frontal Medial Cortex*	16.084	***
*Frontal Operculum Cortex **	19.22	***			
Parahippocampal gyrus, posterior	17.985	***			
Frontal Medial Cortex	17.938	***			

** = significant response only at latency -3*

**** = p < 0.001*

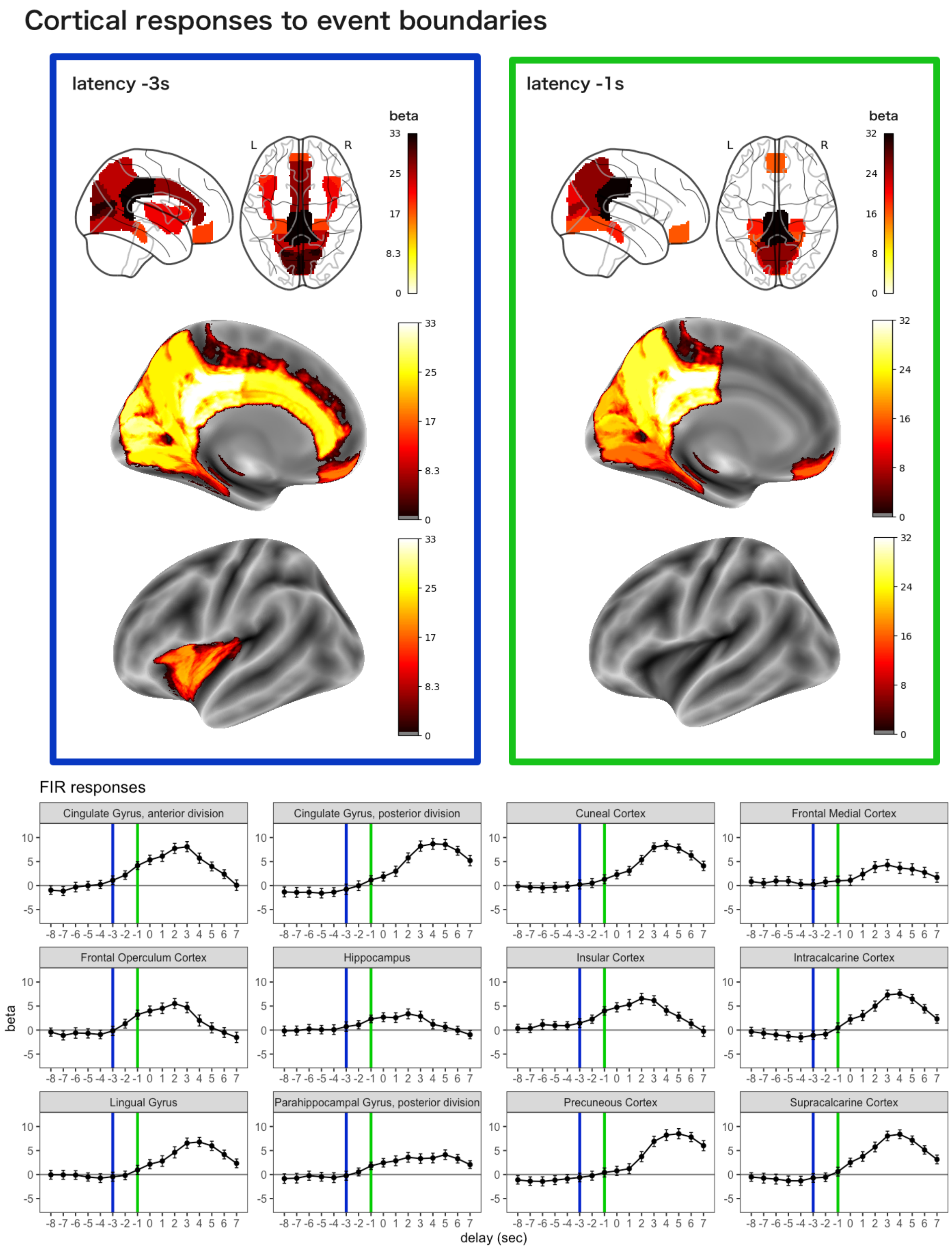
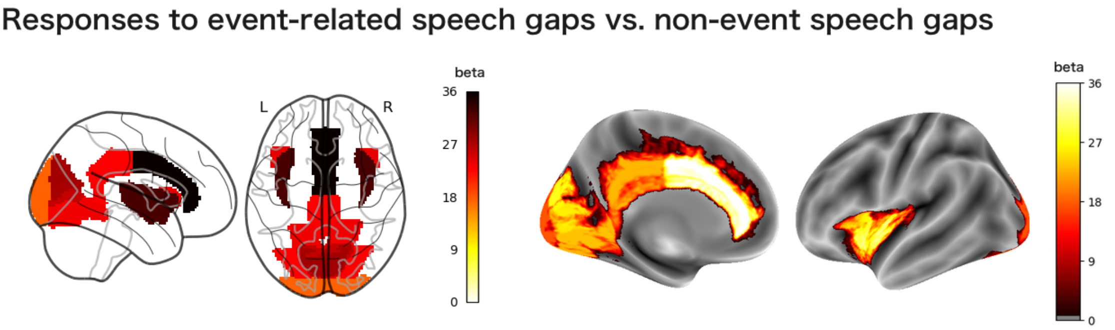


Figure 6: Cortical responses to event boundaries at latencies of -3 sec (blue) and -1 sec (green) relative to the behavioural timing. Most areas respond to both latencies, but ACC, frontal operculum and insula have a significant response only to the earlier latency. Cortical heatmaps and glass brains represent the betas from intact boundaries, thresholded at $p<0.05$ relative to permutations. Note that the symmetry of the activations is due to the atlas used and does not necessarily represent the true activation pattern.

In addition to the shuffled segments the HOA regions were also analysed by controlling with speech gaps, as was done for the hippocampus. The results (Table 4) show that many ROIs remain significant even with this control, most prominently the ACC, insula, and the cuneus (see Figure 7). Additional responses are seen in occipital pole and the occipital fusiform gyrus, which were not significant in the main analysis.

Table 4: Betas from speech gaps related with event boundaries vs non-event speech gaps in the cortical ROIs.

area	beta	p (adj.)
Cingulate Gyrus, anterior division	35.869	***
Insular Cortex	32.239	***
Cuneal Cortex	27.612	***
Supracalcarine Cortex	27.291	***
Intracalcarine Cortex	25.838	***
Frontal Operculum Cortex	23.917	***
Lingual Gyrus	23.426	***
Occipital Fusiform Gyrus	22.829	***
Cingulate Gyrus, posterior division	22.816	***
Occipital Pole	17.496	***



Responses to event-related speech gaps vs. non-event speech gaps

Figure 7: Cortical responses to event-related speech gaps vs non-event speech gaps. Colours represent betas from responses to event gaps, thresholded at adjusted p<0.001 in relation to the non-event responses.

3.4 Response modulation by boundary salience

Modulation of the BOLD responses by boundary salience was investigated with a linear mixed effects model, where the response to each separate boundary at latency -3s was predicted by salience (low N=20, med N=21, and high N=25) and two predictors for low-level auditory features: mean volume and volume difference in a 1 second epoch around the boundary. Boundary and participant ids were added as random effects.

The mean volume correlated somewhat with salience, being smaller for higher salience boundaries (r=-0.265, p=0.018), but volume difference did not (r=0.098, p=0.390).

The hippocampal response was modulated by boundary salience even after adding the two perceptual confound variables (F(1,62)=4.986, p=0.029), but the result did not survive multiple comparison correction. However, several cortical areas exhibited significant modulation, most prominently the cingulate and calcarine gyri and the lingual gyrus (Figure 8). Interestingly, in the PCC the effect of salience modulation increased when the confounds were added, while in all other areas it either decreased or was not affected (Table 5).

Table 5: Response modulation by boundary salience. Results from fitting the mixed-effects model to all ROIs separately, with salience as the independent variable and low-level auditory features (*meanVol*, *volDiff*) as confound predictors. The adjusted p-values are corrected for multiple comparisons with the Holm-Bonferroni method. All areas with corrected p-value for salience less than 0.05 are listed, as well as the results for hippocampus, which do not reach significance after correction. All results are for latency = -3.

	without confounds			**with confounds**		
area	F(1,64)	p	p (adj.)	F(1,62)	p	p (adj.)
Cingulate Gyrus, posterior division	14.236	< 0.001	0.016	16.17	< 0.001	0.008
Cingulate Gyrus, anterior division	20.252	< 0.001	0.001	16.561	< 0.001	0.007
Lingual Gyrus	16.519	< 0.001	0.006	15.289	< 0.001	0.002
Intracalcarine Cortex	16.956	< 0.001	0.005	16.205	< 0.001	0.008
Supracalcarine Cortex	15.777	< 0.001	0.008	15.422	< 0.001	0.01
Frontal Operculum Cortex	15.932	< 0.001	0.008	12.399	0.001	0.035
Subcallosal Cortex	12.302	< 0.001	0.02	12.262	< 0.001	0.021
Hippocampus	6.076	0.016	0.574	4.986	0.029	0.963

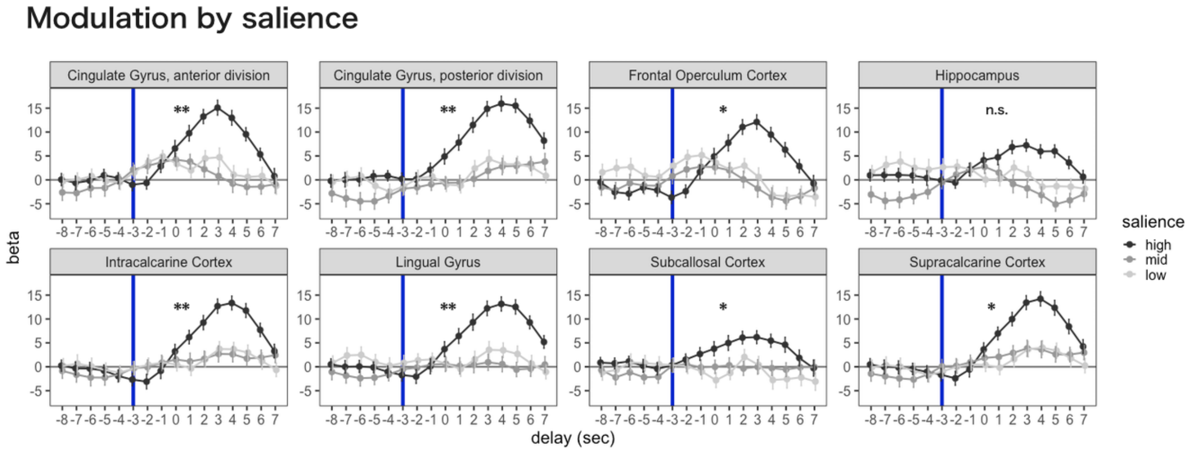


Figure 8: Modulation by salience in areas with significant modulation after the addition of perceptual confounds, as well as in the hippocampus. Asterisks denote the adjusted p-value of salience from the mixed-effects model: ** $p < 0.01$, * $p < 0.05$, n.s. not significant.

The areas of significant modulation reported by Ben-Yakov and Henson (2018) included, in addition to hippocampus, some of the same cortical ROIs, namely PCC and lingual gyrus, but in their data the latter did not remain significant after adding predictors for all perceptual confounds. Yet their list of confounds was much more exhaustive than the one used here – e.g., spectral features of the audio, as well as responses from low-level perceptual areas – so it is likely that some of the effects in the current analysis would disappear with more rigorous controls.

4 Discussion

This study examined brain activations time-locked to behaviourally defined event boundaries in an auditory narrative stimulus, specifically in the hippocampus, as it has been shown to be the crucial link in one-shot learning of episodic memories. The hippocampus was found to be significantly more active in response to the event boundaries, as opposed to either random time points or non-event sentence boundaries.

Additionally, several cortical ROIs were found to respond to event boundaries. Most prominent among these were the PCC, ACC, insula, and the medial occipital areas, with precuneus, vmPFC and PHCG exhibiting activations that did not remain significant when controlled with non-event sentence boundaries – in other words, they are not likely to be driven exclusively by narrative-level event processing. Furthermore, PCC, ACC, frontal operculum, subcallosal cortex, and the medial visual areas showed responses that were modulated by behavioural salience over and above the effect of the most prominent sensory confounds (mean volume and volume difference around the event boundary) – in the hippocampus this effect was not significant.

Familiarity with the story did not significantly affect the hippocampal response in this study. This might be due to the fact that the level of familiarity the participants gained might not have been very high: the 2nd listening participants had only heard the story once before, with at least a month between the first and second hearing. In the study by Ben-Yakov and others, where the familiarity effect was clear, the participants viewed the same short video clips 6 times over two consecutive days (Ben-Yakov et al., 2014). Further, the same study found an increase in the so-called onset response with increasing stimulus familiarity, which is in line with findings of hippocampal activity during memory retrieval and the dual role of HC in event processing (Richmond & Zacks, 2017). In our study the gaps between event offset and onset were essentially non-existent, as the story was continuous, so these responses would necessarily become overlapped and thus mask the possible changes in the offset response.

In the following sections I will discuss the central findings in the context of one-shot learning and episodic memory, as well as event models and segmentation.

4.1 One-shot learning of events

It has long since been known that the hippocampus is required for one-shot learning. By definition, a unique event only happens once, and can only be remembered if it is somehow bound and "written down" right away, and later protected from interference. Much of memory consolidation is believed to happen during sleep (Wixted & Cai, 2013), but it seems obvious that some initial encoding needs to take place much more quickly. An appealing candidate time point for some early encoding activities is the event boundary, which demonstrably affects the structure of memory.

It has been proposed that at event boundaries the hippocampus performs a sharpening function to stabilise the event pattern to facilitate subsequent recall (Richmond & Zacks, 2017). This proposal is supported by evidence of hippocampal BOLD activity peaks at the end of discrete, eventful film clips (Ben-Yakov & Dudai, 2011), in response to pattern shifts in several high-level cortical areas implicated in event model maintenance (Baldassano et al., 2017), and time-locked to behaviourally defined event boundaries in extended film narratives (Ben-Yakov & Henson, 2018). However, all these studies were done utilising dynamic visual or audio-visual stimuli, which leaves open the possibility that the activation is specific to visual experience, and possibly related to visuospatial binding, which the hippocampus is also known to be involved in.

The present study resolves this question by showing hippocampal activation peaks in response to event boundaries in an extended auditory narrative. This means that the hippocampus does indeed follow the abstract, event-level structure of even non-visual stimuli, and is likely engaged in domain-general event segmentation and encoding at event boundaries.

Very recently Cohn-Sheehy and others reported similar findings: the authors examined hippocampal patterns during the listening of an auditory story and reported transient activation peaks at event boundaries (Cohn-Sheehy et al., 2021). Also, an ECoG study by Michelmann and colleagues showed a flow of information from the auditory cortex to the hippocampus at event boundaries in an auditory narrative, which they interpreted as supporting the formation of hippocampal “snapshots” of cortical activity and thus the encoding of the integrated episode (Michelmann et al., 2020). Unfortunately they did not study possible information flows between HC and higher-order cortical areas, such as the PCC.

There is thus a growing body of evidence linking increased hippocampal processing with boundaries in continuous, eventful stimuli. But do these activations reflect a process with a causal role in one-shot learning of events? From studies done with sequences of unconnected and isolated stimuli we know the hippocampal response magnitude to be correlated with subsequent memory performance, which indirectly implies causality, but to my knowledge similar analyses have not been done with extended narrative stimuli. A recent study reported increased hippocampal responses to offsets of events in audio-visual narratives that were more “central”, i.e. more closely connected causally or semantically to other events in the same narrative, and this centrality in turn predicted better recall, but the authors did not test the correlation between response strength and memory directly (Lee & Chen, 2022).

Another way to approach the issue is to consider whether these BOLD peaks might in fact represent hippocampal SWRs, which are known to play a causal role in spatial learning in rodents (Buzsáki, 2015). Recently, Bilkey and Jensen (2021) suggested that human SWRs might provide a neural marker for event boundary detection. However, detecting SWRs in the human hippocampus is quite tricky. Invasive single-unit recordings have been able to reveal activations that might be a human equivalent of SWRs, but their correlation with behaviour was not evaluated (Bragin et al., 1999). Also, invasive methods are not an ideal candidate for extracting neural markers for event boundaries, if we wish to record large numbers of subjects and stimuli. In the macaque brain the occurence of certain subtypes of hippocampal SWRs could be detected as activation peaks in simultaneous fMRI, but most strongly in neocortical areas, especially temporal and parietal cortices, PCC, V1 and the retrosplenial area, and less distinctly in the HC (Ramirez-Villegas et al., 2015). Also, the animals in this study were under anesthesia, so no behavioural correlates for the neural phenomenon exist. In a recent study with humans Liu and colleagues detected a SWR-like increase of high-frequency (140 Hz) power at onsets of cortical rapid replay sequences that source localised to the MTL (Liu et al., 2019). Yet, the design of the study was a sequential learning paradigm, and the replay and putative SWRs took place during rest periods between the memory tasks, which do not necessarily equal event boundaries.

It is thus not clear that SWRs in fact take place in the human hippocampus during event boundaries, or that they might be detected as BOLD peaks in the HC. So while the current hippocampal results support the more general suggestion of some sort of resource-intensive hippocampal activity at event boundaries, which might be linked with sharpening the event representation to enable its encoding (Richmond & Zacks, 2017), we are still a long way from showing its exact cellular underpinnings or a causal role in memory formation.

The gold standard of demonstrating a causal role is, of course, to show that a lesion leads to a behavioural deficit. While amnesia cases provide us with a general-level proof for the necessity of HC for memory encoding, there have unfortunately not been studies where event structure would have been taken explicitly into account, so we do not know whether or not an event segment is in fact the “chunk” of working memory that an amnesic is able to retain. For example, Zuo and colleagues showed recently that a severely amnesic patient is able to retain and integrate information over more than 30 seconds, relying solely on his cortical DMN systems, but even they did not consider event boundaries and their possibly disrupting role in the continuous, narrative stimulus (Zuo et al., 2020).

An alternative to naturally occuring lesions are the developing methods of temporarily modulating neural processing. What would happen, if we were to non-invasively disturb hippocampal processing in humans during event perception? What would be the pattern of deficits, if we were to disturb it at event boundaries, versus at other time points? Would one impair subsequent recall of the event, and the other disrupt the binding of disparate features into the event model, leading to more fragmented and less gist-like recall? Hebscher and others found increased memory performance in response to bursts of theta-band excitatory transcranial magnetic stimulation to cortical "hippocampal networks" (posterior parietal and occipital areas) (Hebscher et al., 2021), but more convincing evidence could be attained by directly targeting the hippocampus with e.g. transcranial focused ultrasound (di Biase et al., 2019) and inhibiting its performance.

4.2 Event models and boundary response

The view that emerged from the mini review of studies investigating boundary-related responses in the brain suggested that the medial parietal responses – in PCC and precuneus – would be most closely connected with high level, modality independent event segmentation, at least with narrative stimuli. This suggestion is partly supported by the current results: PCC does indeed seem to be a very specific responder, whereas precuneus seems more driven by lower-level factors. Precuneus has been implicated in high-level functions such as self-referential thinking, episodic recall, and consciousness (Cavanna & Trimble, 2006), but it is also closely connected with areas involved in sensorimotor, visual and attentive processes (Andrews-Hanna, 2012) and as such probably plays a more lower-level role in event processing than the PCC.

But what does the robust PCC activation mean? Currently, we have very little firm evidence, but some theories can be derived from the functions ascribed to DMN in general and the PCC in particular. For example, a traditional view of DMN links it with self-referential thinking, simulating hypothetical events, episodic recall, affective processing, and spontaneous mind-wandering in general, which typically interferes with task performance (for a comprehensive review see Andrews-Hanna, 2012). In other words, it is thought to support internal mentation, and actively compete for attention with more externally-directed processing networks.

Recently, Stawarczyk, Bezdek and Zacks proposed that the midline DMN core (PCC and amPFC) is largely responsible for the maintenance of the working event model, with particularly PCC being a key area for linking incoming sensory information with knowledge in the LTM (Stawarczyk et al., 2021). They further proposed that PCC activation would be highest when attention is redirected to internal representations and decoupled from the current environment, which would explain its detrimental effect on attention-demanding tasks. Thus the robust activation peak in PCC at narrative event boundaries might indicate a brief shift of attention from external to internal, as e.g. the repercussions of the just-ended event are considered.

This proposition is, however, somewhat contradictory to the much-studied boundary advantage effect in memory, which posits that event boundaries are, on the contrary, time points of *increased* external processing and *less* likely than other moments to elicit mind-wandering (e.g. Baker & Levin, 2015; Faber et al., 2018; Huff et al., 2018). Solving this dilemma would probably require a closer look at how the studies have defined event boundaries, and whether or not the stimuli used were extended and narrative enough to elicit episodic consolidation. Also, the timing of this hypothetical attentional shift might be very specific, with attention snapping back to the external input very swiftly as the new event begins – studies where a boundary advantage has been reported have typically looked at epochs immediately following a boundary, and not those preceding it, which might make all the difference.

It is also possible that activations in PCC specifically or in the DMN in general do not mark attentional shifts, but are just a part of the same sharpening function that is postulated to take place in the hippocampus. Yeshurun and others argue that DMN is in fact not working in opposition to externally-oriented tasks but rather supports them by building and maintaining rich, context-dependent models to facilitate interpretation of incoming information – that it is an “active sense-making network” (Yeshurun et al., 2021). Thus peaks of activity in these areas might serve memory encoding by rapidly replaying the relevant aspects of the just-experienced episode, and help hippocampus to lay out the memory trace.

Also, a recent study by Brandman and others links activation peaks in DMN areas to moments of narrative surprise, and suggests they serve to signal high-level prediction error (Brandman et al., 2021). This is perhaps a complimentary view to the sense-making network proposition, as a surprising twist in the narrative would be likely to trigger an increased need to make sense of the changed situation and revise the old model to fit the new data, which might also lead to increased processing of external stimuli instead of an attentional lapse.

To make sense of the role of PCC activations at event boundaries we would, first of all, need designs which are capable of separating surprise and boundaries. Second, the postulated attentional shifts should somehow be measured, taking care to look at pre- and post-boundary epochs separately. Third, the functional and causal connections between PCC and HC at event boundaries should be investigated, to evaluate the possibility that the transient activation represents rapid neural replay and serves to "feed" the hippocampus. There is some evidence that rapid neural replay does take place in humans at event boundaries (Silva et al., 2019; Sols et al., 2017), but its sources are not easy to pin down with non-invasive methods, though Liu and colleagues suggest that the sources might be neocortical (Liu et al., 2019). An ECoG study similar to that of Michelmann and others (Michelmann et al., 2020), but investigating information flow between HC and high-order DMN areas, would be very informative.R

It is also likely that at least some of the medial visual areas exhibiting boundary-related responses in the current study and in some previous ones (Ezzyat & Davachi, 2011; Speer et al., 2003; Zacks et al., 2006, 2010) take part in populating the event model. For example, neural patterns in the cuneus have been found to distinguish between characters in an interleaved extended film narrative, while patterns in lingual and calcarine gyri distinguished between locations (Milivojevic et al., 2016). However, all of these studies have utilised visual stimuli. Why do we see such robust activations in response to boundaries in a purely auditory narrative?

Generally, visual areas are known to activate in response to visual imagery as well as perception (Pearson & Kosslyn, 2015). A study by Saalasti and others found that while listening to an auditory narrative, participants who shared a higher semantic similarity in their interpretation of the story also showed higher inter-subject correlations in, amongst others, early visual areas, especially the cuneus (Saalasti et al., 2019). It thus seems likely that the activations found in response to event boundaries in our auditory narrative are somehow connected to the visual imagery elicited by the story – but why those areas peak selectively at event boundaries is very much an open question.

Of course, what makes any interpretation of the current (and previous) data even more challenging is the fact that the BOLD response is actually more a measure of the *input* to a certain area than any indication of its output (Logothetis, 2008). It is thus conceivable that the ROIs exhibiting significant BOLD peaks at event boundaries are more prominently targets than sources in the boundary-processing chain.

Also, it should be noted that AG, which is heavily implied in event model construction and upkeep (Baldassano et al., 2017), does not seem to respond with activation peaks to event boundaries. Similarly to PCC it is a part of DMN, though not considered a similar central hub (Andrews-Hanna, 2012). What separates AG and PCC to result in this difference is another open question.

4.3 Prediction error and the boundary response

Of specific interest is the finding in the present study that both ACC and insula respond robustly to boundaries, but *only* in the earlier, -3s latency condition. In other words, their BOLD peaks precede those of PCC and the medial occipital areas by a second or more. ACC and the anterior insula are anatomically and functionally connected, and are proposed to form a "salience network", which evaluates and tags important internal and external events to help direct attentional resources, possibly dynamically switching between DMN and the more task-oriented central executive network (Menon & Uddin, 2010). Also, ACC specifically has long been though to be involved in conflict monitoring and decision making (Botvinick, 2007), and in fact the original formulation of EST predicted that ACC should be prominently active at event boundaries, signalling prediction error (Zacks & Swallow, 2007). This was, however, not systematically found, and more recent formulations of the theory have dropped the subject (Richmond & Zacks, 2017; Zacks, 2020). Note though that the signalling of uncertainty is not a unitary thing, so the absence of ACC activations as such does not mean that there would be no processing of prediction error (see e.g. Bach & Dolan, 2012).

Previously ACC responses to boundaries have been found very inconsistently, with no clear connection to stimulus or boundary type, or any other design feature. The findings in the present study, however, are quite clear, and even more convincing when we consider the specificity of the activations in ACC and insula to event boundaries compared with other sentence boundaries. What could explain this discrepancy? One possible explanation might be that in the current study the boundaries with the earlier latency (-3s) are really quite early, and tend to coincide with the speech gaps between the events. It might be that in previous studies the boundaries have been defined in a way which does not capture these very early, perhaps preparatory, activations. It would be very interesting to see whether similar activations precede boundaries in other datasets as well, and if they do, whether they contribute to triggering activations in other areas, which would fit well with assumed role of the salience network in dynamic switching (Menon & Uddin, 2010).

To determine how the brain manages the online segmentation of experience into useful event models, we should gain a clearer understanding of its triggers. It is likely that while segmentation is at least partly driven by external cues, which might give rise to prediction error, it also relies heavily on internal expectations in the sense that a complete event schema also contains information on when the event is likely to end. This overall logic could be captured with a measure of uncertainty, which, unlike prediction error, does not require surprise, or a mismatch between expectation and outcome. So far very little empirical research has concentrated on this question, even though theoretical models of segmentation abound. To base these theories on a firmer ground it would be prudent to empirically test the relative effects of surprise and uncertainty on boundary detection, as well as the neural correlates thereof.

4.4 Caveats

The major caveats of this study relate to the fact that our stimulus is a naturalistic one. Several low-level features of the narrative, e.g. the rhythm of speech or prosody, could not be controlled for, and thus it is an open question how these features guide the segmentation of the story, and what is their share in the neural signals observed.

Some elements could be controlled for by controlling against random non-event sentence boundaries, but we know that at least the durations of speech gaps differed between events and non-events, and the mean volume around event boundary related speech gaps correlated rather strongly with perceived salience. So the gap lengths certainly affect segmentation, and the role of prosody was not even examined. The gravity of the issue can be appreciated in light of the finding that purely non-linguistic cues enable a coarse-level segmentation of an audio drama, i.e., people who do not understand the language used are nevertheless able to spot nearly all major event boundaries (Papenmeier et al., 2019). Yet, we did not observe significant responses to event boundaries in any auditory processing areas, so in light of that, as well as results from earlier studies utilising very different stimuli, we may conclude that mostly the neural signals reported here relate to high-level, domain-general segmentation processes.

Another caveat to consider is the same that was also noted by Ben-Yakov and Henson (2018): namely, the operationalisation of perceived salience as the number of observers marking a certain boundary. It is by no means clear that this in fact represents *subjectively* perceived salience, i.e. the experience of an exceptionally clear and unambiguous boundary. Especially as I made the choice to allow rather long clusters when forming the combined boundaries, some intrinsically rather diffuse shifts might have gained high salience; though generally salience and cluster length did not correlate, i.e., most of the high salience boundaries were also rather concise. A related issue is the fact that the confounds used to control the effect of salience on response magnitude were not as extensive as those extracted by Ben-Yakov and Henson (2018) – therefore it cannot be ruled out that the modulations showing as significant in these analyses might in fact not remain so with more stringent controls. All results concerning salience modulation are thus rather inconclusive: cortical areas might be responding to uncontrolled confounds more than to salience, and the hippocampal response might be significant if salience was measured directly, by subjective evaluation.

The final caveat pertains to the selected method of extracting the hippocampal ROI: due to limited resources this was not done with the state-of-the-art technique of individually delineating each participant’s hippocampus, but instead by masking in MNI space with the Harvard-Oxford subcortical atlas. This is known to be a suboptimal method (Squire et al., 2004), which can seriously understate group-level effects. Fortunately, this does not invalidate the hippocampal results, but does act as a possibly considerable source of noise, which might lead to underestimation of the significance of the findings.

4.5 Future directions

The study of human memory systems – as well as other areas of cognition – benefits much from the use of relatively naturalistic, temporally extended stimuli (see e.g. Jääskeläinen et al., 2021; Lee et al., 2020; Sonkusare et al., 2019). The human capacity for one-shot learning of events and memory items is rather astonishing, and even more so when one considers that in realistic situations the schedule of building, encoding, and abandoning event models needs to be decided on the spot, based on ambiguous external cues and imprecise internal predictions. The use of narratives helps uncover how the brain manages this. It is possibly even misleading to generalise from the highly artificial memory tasks used in traditional lab-based research to the functioning of our everyday memory capacity (for a general discussion, see Hasson et al., 2020). After all, the problems people face with memory are related to everyday life and its demands, and to target those problems we need to understand their basis – even if they are complex and not amenable to convenient theoretical dissection.

That said, it is by no means necessary for all memory or segmentation research to focus exclusively on narratives: it is just as important to continue looking into the full stack of processing levels and their concomitant segmentation windows. But what is required is a more rigorous framework for placing the gained results – e.g., neural responses – into their relevant contexts, defined by the stimulus and experimental setting used. In other words, segmentation is part of the general online processing of experience, and it should be interpreted in the same hierarchical framework, for example the one proposed by Hasson and others (Hasson et al., 2015). It also needs to be recognised that the human propensity for behavioural segmentation is likely to be at least somewhat flexible and context-dependent: given a highly simplified stimulus it will deliver boundaries, but their “meaning” is not the same as with a more complex stimulus, and the neural representations are thus also unlikely to overlap, which is evident in the fMRI literature of segmentation reviewed in this work.

Also, using more complex naturalistic stimuli does not mean abandoning experimental manipulations. It is quite possible to create experimental designs where the stimuli are naturalistic and engaging, and still subtly manipulated. A very good example of this is a recent study by Cohn-Sheehy and others (2021), where the authors wrote short stories with either connected or disconnected secondary storylines to investigate whether this narrative coherence is represented in the hippocampus (it was). Another clever design, which did not even require manipulating the stimulus, was devised by Lee and Chen (2022): they transformed linear narratives into networks of events, based on semantic and causal connections, and analysed boundary responses in terms of this network centrality. Future work in this area would be greatly aided by collaboration with the professionals from the field of narrative storytelling.

4.6 Conclusion

The present study adds to the growing body of evidence garnered from humans engaged in relatively naturalistic event perception, and shows that the hippocampus responds with transient activation peaks to behaviourally defined event boundaries. As the stimulus used was an auditory narrative, the possibility that this response is dependent on visuospatial processing can be excluded. The results support the general idea that the hippocampus performs a sharpening function at event boundaries, which stabilises the event pattern and enables subsequent recall, as proposed by Richmond and Zacks (2017).

Futhermore, the current study found significant cortical boundary-related activations mainly in posterior medial ROIs, though extending to vmPFC, ACC and insula. This supports the notion that especially PCC is an important area for high-level event models and segmentation, and raises the possibility that some sort of general attention-shifting takes place at event boundaries. It is not clear why the medial visual areas also seem sensitive to boundaries in an auditory narrative – similar findings have not been reported before.

It is likely that the hippocampal and cortical activations are connected, either directly or by being triggered by the same underlying neural process, but the causal connections within the brain and with behaviour need to be elucidated in future studies.

5 References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. 2014. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*, 14. https://doi.org/10.3389/fninf.2014.00014

Ambrose, R. E., Pfeiffer, B. E., & Foster, D. J. 2016. Reverse Replay of Hippocampal Place Cells Is Uniquely Modulated by Changing Reward. *Neuron*, *91*5, 1124–1136. https://doi.org/10.1016/j.neuron.2016.07.047

Andrews-Hanna, J. R. 2012. The Brain's Default Network and Its Adaptive Role in Internal Mentation. *The Neuroscientist*, *18*3, 251–270. https://doi.org/10.1177/1073858411403316

Asano, S., Mihara, B., Kirino, T., & Sugishita, M. 2004. Anatomical constraints on visualization of the human hippocampus using echo-planar imaging. *Neuroradiology*, *46*7. https://doi.org/10.1007/s00234-004-1225-z

Bach, D. R., & Dolan, R. J. 2012. Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience*, *13*8, 572–586. https://doi.org/10.1038/nrn3289

Baddeley, A. 2000. The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*11, 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A., & Wilson, B. A. 2002. Prose recall and amnesia: implications for the structure of working memory. *Neuropsychologia*, *40*10, 1737–1743. https://doi.org/10.1016/S0028-3932(01)00146-4

Baker, L. J., & Levin, D. T. 2015. The role of relational triggers in event perception. *Cognition*, *136*, 14–29. https://doi.org/10.1016/j.cognition.2014.11.030

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. 2017. Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, *95*3, 709-721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Baldassano, C., Hasson, U., & Norman, K. A. 2018. Representation of Real-World Event Schemas during Narrative Perception. *The Journal of Neuroscience*, *38*45, 9689–9699. https://doi.org/10.1523/JNEUROSCI.0251-18.2018

Baldwin, D. A., & Kosie, J. E. 2021. How Does the Mind Render Streaming Experience as Events? *Topics in Cognitive Science*, *13*1, 79–105. https://doi.org/10.1111/tops.12502

Bates, D., Mächler, M., Bolker, B., & Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*1. https://doi.org/10.18637/jss.v067.i01

Ben-Yakov, A., & Dudai, Y. 2011. Constructing Realistic Engrams: Poststimulus Activity of Hippocampus and Dorsal Striatum Predicts Subsequent Episodic Memory. *Journal of Neuroscience*, *31*24, 9032–9042. https://doi.org/10.1523/JNEUROSCI.0702-11.2011

Ben-Yakov, A., Eshel, N., & Dudai, Y. 2013. Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. *Journal of Experimental Psychology: General*, *142*4, 1255–1263. https://doi.org/10.1037/a0033558

Ben-Yakov, A., & Henson, R. N. 2018. The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. *The Journal of Neuroscience*, *38*47, 10057–10068. https://doi.org/10.1523/JNEUROSCI.0524-18.2018

Ben-Yakov, A., Rubinson, M., & Dudai, Y. 2014. Shifting Gears in Hippocampus: Temporal Dissociation between Familiarity and Novelty Signatures in a Single Event. *Journal of Neuroscience*, *34*39, 12973–12981. https://doi.org/10.1523/JNEUROSCI.1892-14.2014

Bilkey, D. K., & Jensen, C. 2021. Neural Markers of Event Boundaries. *Topics in Cognitive Science*, *13*1, 128–141. https://doi.org/10.1111/tops.12470

Bird, C. M., Keidel, J. L., Ing, L. P., Horner, A. J., & Burgess, N. 2015. Consolidation of Complex Events via Reinstatement in Posterior Cingulate Cortex. *Journal of Neuroscience*, *35*43, 14426–14434. https://doi.org/10.1523/JNEUROSCI.1774-15.2015

Bird, Chris M. 2020. How do we remember events? *Current Opinion in Behavioral Sciences*, *32*, 120–125. https://doi.org/10.1016/j.cobeha.2020.01.020

Botvinick, M. M. 2007. Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience*, *7*4, 356–366. https://doi.org/10.3758/CABN.7.4.356

Bragin, A., Engel, J., Wilson, C. L., Fried, I., & Buzs□ki, G. 1999. High-frequency oscillations in human brain. *Hippocampus*, *9*2, 137–142. https://doi.org/10.1002/(SICI)1098-1063(1999)9:2<137::AID-HIPO5>3.0.CO;2-0

Brandman, T., Malach, R., & Simony, E. 2021. The surprising role of the default mode network in naturalistic perception. *Communications Biology*, *4*1, 79. https://doi.org/10.1038/s42003-020-01602-z

Buzsáki, G. 2015. Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus*, *25*10, 1073–1188. https://doi.org/10.1002/hipo.22488

Cavanna, A. E., & Trimble, M. R. 2006. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, *129*3, 564–583. https://doi.org/10.1093/brain/awl004

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. 2017. Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*1, 115–125. https://doi.org/10.1038/nn.4450

Cohen, N. J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., & Nash, C. 1999. *Hippocampal system and declarative (relational) memory: Summarizing the data from functional neuroimaging studies*. 16.

Cohn-Sheehy, B. I., Delarazan, A. I., Reagh, Z. M., Crivelli-Decker, J. E., Kim, K., Barnett, A. J., Zacks, J. M., & Ranganath, C. 2020. *Building bridges: the hippocampus constructs narrative memories across distant events* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.12.11.422162

Cohn-Sheehy, B. I., Delarazan, A. I., Reagh, Z. M., Crivelli-Decker, J. E., Kim, K., Barnett, A. J., Zacks, J. M., & Ranganath, C. 2021. The hippocampus constructs narrative memories across distant events. *Current Biology*, *31*22, 4935-4945.e7. https://doi.org/10.1016/j.cub.2021.09.013

Danker, J. F., & Anderson, J. R. 2010. The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin*, *136*1, 87–102. https://doi.org/10.1037/a0017937

Davachi, L., & Danker, J. 2013. *Cognitive Neuroscience of Episodic Memory*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199988693.013.0018 di Biase, L., Falato, E., & Di Lazzaro, V. 2019. Transcranial Focused Ultrasound (tFUS) and Transcranial Unfocused Ultrasound (tUS) Neuromodulation: From Theoretical Principles to Stimulation Practices. *Frontiers in Neurology*, *10*, 549. https://doi.org/10.3389/fneur.2019.00549

Ego-Stengel, V., & Wilson, M. A. 2009. Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus*, NA-NA. https://doi.org/10.1002/hipo.20707

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

Faber, M., Radvansky, G. A., & D'Mello, S. K. 2018. Driven to distraction: A lack of change gives rise to mind wandering. *Cognition*, *173*, 133–137. https://doi.org/10.1016/j.cognition.2018.01.007

Franklin, N. T., Norman, K. A., Ranganath, C., Zacks, J. M., & Gershman, S. J. 2020. *Structured Event Memory: A Neuro-Symbolic Model of Event Cognition*. 35.

Gernsbacher, M. A. 1985. Surface information loss in comprehension. *Cognitive Psychology*, *17*3, 324–363. https://doi.org/10.1016/0010-0285(85)90012-X

Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G., & Zugaro, M. B. 2009. Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neuroscience*, *12*10, 1222–1223. https://doi.org/10.1038/nn.2384

Greicius, M. D., Krasnow, B., Boyett-Anderson, J. M., Eliez, S., Schatzberg, A. F., Reiss, A. L., & Menon, V. 2003. Regional analysis of hippocampal activation during memory encoding and retrieval: fMRI study. *Hippocampus*, *13*1, 164–174. https://doi.org/10.1002/hipo.10064

Hakonen, M., Ikäheimonen, A., Hultèn, A., Kauttonen, J., Koskinen, M., Lin, F.-H., Lowe, A., Sams, M., & Jääskelainen, I. P. 2020. *Processing of a spoken narrative in the human brain is shaped by family cultural background* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.05.11.083931

Hari, R., & Puce, A. 2017. *MEG-EEG primer*. Oxford University Press.

Hasson, U., Chen, J., & Honey, C. J. 2015. Hierarchical process memory: memory as an integral component of information processing. *Trends in Cognitive Sciences*, *19*6, 304–313. https://doi.org/10.1016/j.tics.2015.04.006

Hasson, U., Nastase, S. A., & Goldstein, A. 2020. Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks. *Neuron*, *105*3, 416–434. https://doi.org/10.1016/j.neuron.2019.12.002

Hebscher, M., Kragel, J. E., Kahnt, T., & Voss, J. L. 2021. Enhanced reinstatement of naturalistic event memories due to hippocampal-network-targeted stimulation. *Current Biology*, *31*7, 1428-1437.e5. https://doi.org/10.1016/j.cub.2021.01.027

Henson, R. 2005. A Mini-Review of fMRI Studies of Human Medial Temporal Lobe Activity Associated with Recognition Memory. *The Quarterly Journal of Experimental Psychology Section B*, *58*3–4b, 340–360. https://doi.org/10.1080/02724990444000113

Holm, S. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*2, 65–70. JSTOR.

Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J., & Burgess, N. 2015. Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications*, *6*1, 7462. https://doi.org/10.1038/ncomms8462

Hsu, Y.-C., Chu, Y.-H., Tsai, S.-Y., Kuo, W.-J., Chang, C.-Y., & Lin, F.-H. 2017. Simultaneous multi-slice inverse imaging of the human brain. *Scientific Reports*, *7*1, 17019. https://doi.org/10.1038/s41598-017-16976-0

Huettel, S. A., Song, A. W., & McCarthy, G. 2014. *Functional Magnetic Resonance Imaging* 3rd ed. Sinauer.

Huff, M., Maurer, A. E., Brich, I., Pagenkopf, A., Wickelmaier, F., & Papenmeier, F. 2018. Construction and updating of event models in auditory event processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*2, 307–320. https://doi.org/10.1037/xlm0000482

Huff, M., Meitz, T. G. K., & Papenmeier, F. 2014. Changes in situation models modulate processes of event perception in audiovisual narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*5, 1377–1388. https://doi.org/10.1037/a0036780

Huff, M., Papenmeier, F., & Zacks, J. M. 2012. Visual target detection is impaired at event boundaries. *Visual Cognition*, *20*7, 848–864. https://doi.org/10.1080/13506285.2012.705359

Jääskeläinen, I. P., Sams, M., Glerean, E., & Ahveninen, J. 2021. Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage*, *224*, 117445. https://doi.org/10.1016/j.neuroimage.2020.117445

Jadhav, S. P., Kemere, C., German, P. W., & Frank, L. M. 2012. Awake Hippocampal Sharp-Wave Ripples Support Spatial Memory. *Science*, *336*6087, 1454–1458. https://doi.org/10.1126/science.1217230

Khemlani, S. S., Harrison, A. M., & Trafton, J. G. 2015. Episodes, events, and models. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00590

Kirwan, C. B., & Stark, C. E. L. 2004. Medial temporal lobe activation during encoding and retrieval of novel face-name pairs. *Hippocampus*, *14*7, 919–930. https://doi.org/10.1002/hipo.20014

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*13. https://doi.org/10.18637/jss.v082.i13

Lawrence, Z., & Peterson, D. 2016. Mentally walking through doorways causes forgetting: The location updating effect and imagination. *Memory*, *24*1, 12–20. https://doi.org/10.1080/09658211.2014.980429

Lee, H., Bellana, B., & Chen, J. 2020. What can narratives tell us about the neural bases of human memory? *Current Opinion in Behavioral Sciences*, *32*, 111–119. https://doi.org/10.1016/j.cobeha.2020.02.007

Lee, H., & Chen, J. 2022. Predicting memory from the network structure of naturalistic events. *Nature Communications*, *13*1, 4235. https://doi.org/10.1038/s41467-022-31965-2

Leuner, B., & Gould, E. 2010. Structural Plasticity and Hippocampal Function. *Annual Review of Psychology*, *61*1, 111–140. https://doi.org/10.1146/annurev.psych.093008.100359

Liu, Y., Dolan, R. J., Kurth-Nelson, Z., & Behrens, T. E. J. 2019. Human Replay Spontaneously Reorganizes Experience. *Cell*, *178*3, 640-652.e14. https://doi.org/10.1016/j.cell.2019.06.012

Logothetis, N. K. 2008. What we can do and what we cannot do with fMRI. *Nature*, *453*7197, 869–878. https://doi.org/10.1038/nature06976

Magliano, J. P., Radvansky, G. A., Forsythe, J. C., & Copeland, D. E. 2014. Event segmentation during first-person continuous events. *Journal of Cognitive Psychology*, *26*6, 649–661. https://doi.org/10.1080/20445911.2014.930042

McDermott, K. B., Szpunar, K. K., & Christ, S. E. 2009. Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia*, *47*11, 2290–2298. https://doi.org/10.1016/j.neuropsychologia.2008.12.025

Menon, V., & Uddin, L. Q. 2010. Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, *214*5–6, 655–667. https://doi.org/10.1007/s00429-010-0262-0

Michelmann, S., Price, A. R., Aubrey, B., Doyle, W. K., Friedman, D., Dugan, P. C., Devinsky, O., Devore, S., Flinker, A., Hasson, U., & Norman, K. A. 2020. *Moment-by-moment tracking of naturalistic learning and its underlying hippocampo-cortical interactions* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.12.09.416438

Milivojevic, B., Varadinov, M., Vicente Grabovetsky, A., Collin, S. H. P., & Doeller, C. F. 2016. Coding of Event Nodes and Narrative Context in the Hippocampus. *The Journal of Neuroscience*, *36*49, 12412–12424. https://doi.org/10.1523/JNEUROSCI.2889-15.2016

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. 2016. Episodic Memory and Beyond: The Hippocampus and Neocortex in Transformation. *Annual Review of Psychology*, *67*1, 105–134. https://doi.org/10.1146/annurev-psych-113011-143733

Nastase, S. A., Goldstein, A., & Hasson, U. 2020. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*, 117254. https://doi.org/10.1016/j.neuroimage.2020.117254

Newtson, D. 1973. Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*1, 28–38. https://doi.org/10.1037/h0035584

Newtson, D., & Engquist, G. 1976. The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, *12*5, 436–450. https://doi.org/10.1016/0022-1031(76)90076-7

Newtson, D., Engquist, G. A., & Bois, J. 1977. The objective basis of behavior units. *Journal of Personality and Social Psychology*, *35*12, 847–862. https://doi.org/10.1037/0022-3514.35.12.847

Olsen, R. K., Moses, S. N., Riggs, L., & Ryan, J. D. 2012. The hippocampus supports multiple cognitive processes through relational binding and comparison. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00146

Paller, K. A., & Mayes, R. 1987. *Neural correlates of encoding in an incidental learning paradigm*. 12.

Paller, K. A., & Wagner, A. D. 2002. Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, *6*2, 93–102. https://doi.org/10.1016/S1364-6613(00)01845-3

Pamilo, S., Malinen, S., Hotta, J., & Seppä, M. 2015. A correlation-based method for extracting subject-specific components and artifacts from group-fMRI data. *The European Journal of Neuroscience*, *42*9, 2726–2741. https://doi.org/10.1111/ejn.13034

Papenmeier, F., Maurer, A. E., & Huff, M. 2019. Linguistic Information in Auditory Dynamic Events Contributes to the Detection of Fine, Not Coarse Event Boundaries. *Advances in Cognitive Psychology*, *15*1, 30–40. https://doi.org/10.5709/acp-0254-9

Pearson, J., & Kosslyn, S. M. 2015. The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences*, *112*33, 10089–10092. https://doi.org/10.1073/pnas.1504933112

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*1, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pettijohn, K. A., & Radvansky, G. A. 2016a. Walking through doorways causes forgetting: environmental effects. *Journal of Cognitive Psychology*, *28*3, 329–340. https://doi.org/10.1080/20445911.2015.1123712

Pettijohn, K. A., & Radvansky, G. A. 2016b. Narrative event boundaries, reading times, and expectation. *Memory & Cognition*, *44*7, 1064–1075. https://doi.org/10.3758/s13421-016-0619-6

Pettijohn, K. A., & Radvansky, G. A. 2018. Walking through doorways causes forgetting: recall. *Memory*, *26*10, 1430–1435. https://doi.org/10.1080/09658211.2018.1489555

Pfeiffer, B. E., & Foster, D. J. 2013. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, *497*7447, 74–79. https://doi.org/10.1038/nature12112

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Radvansky, G. A., & Copeland, D. E. 2006. Walking through doorways causes forgetting: Situation models and experienced space. *Memory & Cognition*, *34*5, 1150–1156. https://doi.org/10.3758/BF03193261

Radvansky, G. A., Krawietz, S. A., & Tamplin, A. K. 2011. Walking through Doorways Causes Forgetting: Further Explorations. *Quarterly Journal of Experimental Psychology*, *64*8, 1632–1645. https://doi.org/10.1080/17470218.2011.571267

Ramirez-Villegas, J. F., Logothetis, N. K., & Besserve, M. 2015. Diversity of sharp-wave–ripple LFP signatures reveals differentiated brain-wide dynamical events. *Proceedings of the National Academy of Sciences*, *112*46, E6379–E6387. https://doi.org/10.1073/pnas.1518257112

Ranganath, C., & Hsieh, L.-T. 2016. The hippocampus: a special place for time. *Annals of the New York Academy of Sciences*, *1369*1, 93–110. https://doi.org/10.1111/nyas.13043

Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. 2004. Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, *42*1, 2–13. https://doi.org/10.1016/j.neuropsychologia.2003.07.006

Reagh, Z. M., & Ranganath, C. 2021. *A cortico-hippocampal scaffold for representing and recalling lifelike events* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.04.16.439894

Reas, E. T., & Brewer, J. B. 2013. Effortful retrieval reduces hippocampal activity and impairs incidental encoding. *Hippocampus*, *23*5, 367–379. https://doi.org/10.1002/hipo.22096

Rempel-Clower, N. L., Zola, S. M., Squire, L. R., & Amaral, D. G. 1996. Three Cases of Enduring Memory Impairment after Bilateral Damage Limited to the Hippocampal Formation. *The Journal of Neuroscience*, *16*16, 5233–5255. https://doi.org/10.1523/JNEUROSCI.16-16-05233.1996

Richmond, L. L., & Zacks, J. M. 2017. Constructing Experience: Event Models from Perception to Action. *Trends in Cognitive Sciences*, *21*12, 962–980. https://doi.org/10.1016/j.tics.2017.08.005

Rissman, J., Gazzaley, A., & D'Esposito, M. 2004. Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*, *23*2, 752–763. https://doi.org/10.1016/j.neuroimage.2004.06.035

Ritchey, M., Libby, L. A., & Ranganath, C. 2015. Cortico-hippocampal systems involved in memory and cognition. In *Progress in Brain Research* Vol. 219, pp. 45–64. Elsevier. https://doi.org/10.1016/bs.pbr.2015.04.001

Ritchey, M., Wing, E. A., LaBar, K. S., & Cabeza, R. 2013. Neural Similarity Between Encoding and Retrieval is Related to Memory Via Hippocampal Interactions. *Cerebral Cortex*, *23*12, 2818–2828. https://doi.org/10.1093/cercor/bhs258

RStudio Team. 2019. *RStudio: Integrated Development Environment for R*. RStudio, Inc. http://www.rstudio.com/

Saalasti, S., Alho, J., Bar, M., Glerean, E., Honkela, T., Kauppila, M., Sams, M., & Jääskeläinen, I. P. 2019. Inferior parietal lobule and early visual areas support elicitation of individualized meanings during narrative listening. *Brain and Behavior*, *9*5, e01288. https://doi.org/10.1002/brb3.1288

Sasmita, K., & Swallow, K. M. 2021. *Measuring Events: An Investigation into the Stability of Event Segmentation Patterns Across Groups* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/2tu6c

Seabold, S., & Perktold, J. 2010. *Statsmodels: Econometric and Statistical Modeling with Python*. 92–96. https://doi.org/10.25080/Majora-92bf1922-011

Seel, S. V., Easton, A., McGregor, A., Buckley, M. G., & Eacott, M. J. 2019. Walking through doorways differentially affects recall and familiarity. *British Journal of Psychology*, *110*1, 173–184. https://doi.org/10.1111/bjop.12343

Shin, Y. S., & DuBrow, S. 2021. Structuring Memory Through Inference-Based Event Segmentation. *Topics in Cognitive Science*, *13*1, 106–127. https://doi.org/10.1111/tops.12505

Silva, M., Baldassano, C., & Fuentemilla, L. 2019. Rapid Memory Reactivation at Movie Event Boundaries Promotes Episodic Encoding. *The Journal of Neuroscience*, *39*43, 8538–8548. https://doi.org/10.1523/JNEUROSCI.0360-19.2019

Singer, A. C., Carr, M. F., Karlsson, M. P., & Frank, L. M. 2013. Hippocampal SWR Activity Predicts Correct Decisions during the Initial Learning of an Alternation Task. *Neuron*, *77*6, 1163–1173. https://doi.org/10.1016/j.neuron.2013.01.027

Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. 2017. Event Boundaries Trigger Rapid Memory Reinstatement of the Prior Events to Promote Their Representation in Long-Term Memory. *Current Biology*, *27*22, 3499-3504.e4. https://doi.org/10.1016/j.cub.2017.09.057

Sonkusare, S., Breakspear, M., & Guo, C. 2019. Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends in Cognitive Sciences*, *23*8, 699–714. https://doi.org/10.1016/j.tics.2019.05.004

Spaniol, J., Davidson, P. S. R., Kim, A. S. N., Han, H., Moscovitch, M., & Grady, C. L. 2009. Event-related fMRI studies of episodic encoding and retrieval: Meta-analyses using activation likelihood estimation. *Neuropsychologia*, *47*8–9, 1765–1779. https://doi.org/10.1016/j.neuropsychologia.2009.02.028

Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. 2009. Reading Stories Activates Neural Representations of Visual and Motor Experiences. *Psychological Science*, *20*8, 989–999. https://doi.org/10.1111/j.1467-9280.2009.02397.x

Speer, N. K., Swallow, K. M., & Zacks, J. M. 2003. Activation of human motion processing areas during event perception. *Cognitive, Affective, & Behavioral Neuroscience*, *3*4, 335–345. https://doi.org/10.3758/CABN.3.4.335

Speer, N. K., & Zacks, J. M. 2005. Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts☆. *Journal of Memory and Language*, *53*1, 125–140. https://doi.org/10.1016/j.jml.2005.02.009

Speer, N. K., Zacks, J. M., & Reynolds, J. R. 2007. Human Brain Activity Time-Locked to Narrative Event Boundaries. *Psychological Science*, *18*5, 449–455. https://doi.org/10.1111/j.1467-9280.2007.01920.x

Squire, L. R., Stark, C. E. L., & Clark, R. E. 2004. THE MEDIAL TEMPORAL LOBE. *Annual Review of Neuroscience*, *27*1, 279–306. https://doi.org/10.1146/annurev.neuro.27.070203.144130

Squire, L. R., & Wixted, J. T. 2011. The Cognitive Neuroscience of Human Memory Since H.M. *Annual Review of Neuroscience*, *34*1, 259–288. https://doi.org/10.1146/annurev-neuro-061010-113720

Squire, L. R., & Zola-Morgan, S. 1988. Memory: brain systems and behavior. *Trends in Neurosciences*, *11*4, 170–175. https://doi.org/10.1016/0166-2236(88)90144-0

Staresina, B. P., Cooper, E., & Henson, R. N. 2013. Reversible Information Flow across the Medial Temporal Lobe: The Hippocampus Links Cortical Modules during Memory Retrieval. *Journal of Neuroscience*, *33*35, 14184–14192. https://doi.org/10.1523/JNEUROSCI.1987-13.2013

Staresina, Bernhard P., & Davachi, L. 2009. Mind the Gap: Binding Experiences across Space and Time in the Human Hippocampus. *Neuron*, *63*2, 267–276. https://doi.org/10.1016/j.neuron.2009.06.024

Stark, C. E. L., & Okado, Y. 2003. Making Memories without Trying: Medial Temporal Lobe Activity Associated with Incidental Memory Formation during Recognition. *The Journal of Neuroscience*, *23*17, 6748–6753. https://doi.org/10.1523/JNEUROSCI.23-17-06748.2003

Stawarczyk, D., Bezdek, M. A., & Zacks, J. M. 2021. Event Representations and Predictive Processing: The Role of the Midline Default Network Core. *Topics in Cognitive Science*, *13*1, 164–186. https://doi.org/10.1111/tops.12450

St-Laurent, M., Abdi, H., & Buchsbaum, B. R. 2015. Distributed Patterns of Reactivation Predict Vividness of Recollection. *Journal of Cognitive Neuroscience*, *27*10, 2000–2018. https://doi.org/10.1162/jocn_a_00839

Swallow, K. M., Barch, D. M., Head, D., Maley, C. J., Holder, D., & Zacks, J. M. 2011. Changes in Events Alter How People Remember Recent Information. *Journal of Cognitive Neuroscience*, *23*5, 1052–1064. https://doi.org/10.1162/jocn.2010.21524

Swallow, K. M., Zacks, J. M., & Abrams, R. A. 2009. Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*2, 236–257. https://doi.org/10.1037/a0015631

Torchiano, M. 2020. *effsize: Efficient Effect Size Computation*. https://doi.org/10.5281/zenodo.1480624

Tulving, E. 1972. Episodic and semantic memory. *Organization of Memory*, *1*, 381–403.

Tulving, E. 2002. Episodic Memory: From Mind to Brain. *Annual Review of Psychology*, *53*1, 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Westfall, J., Nichols, T. E., & Yarkoni, T. 2017. Fixing the stimulus-as-fixed-effect fallacy in task fMRI. *Wellcome Open Research*, *1*, 23. https://doi.org/10.12688/wellcomeopenres.10298.2

Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., & Kircher, T. 2009. Neural correlates of narrative shifts during auditory story comprehension. *NeuroImage*, *47*1, 360–366. https://doi.org/10.1016/j.neuroimage.2009.04.037

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Winocur, G., & Moscovitch, M. 2011. Memory Transformation and Systems Consolidation. *Journal of the International Neuropsychological Society*, *17*05, 766–780. https://doi.org/10.1017/S1355617711000683

Wixted, J., & Cai, D. J. 2013. *Memory Consolidation*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199988693.013.0021

Yeshurun, Y., Nguyen, M., & Hasson, U. 2021. The default mode network: where the idiosyncratic self meets the shared social world. *Nature Reviews Neuroscience*, *22*3, 181–192. https://doi.org/10.1038/s41583-020-00420-w

Zacks, J. M. 2020. *Event Perception and Memory*. 29.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. 2001. Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*6, 651–655. https://doi.org/10.1038/88486

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., Nicole K. Speer, Swallow, Khena M., & Maley, Corey J. 2010. The brain's cutting-room floor: segmentation of narrative cinema. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00168

Zacks, J. M., Speer, N. K., & Reynolds, J. R. 2009. Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*2, 307–327. https://doi.org/10.1037/a0015305

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., & Swallow, K. M. 2007. Event Segmentation. *Current Directions in Psychological Science*, *16*2, 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x

Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. 2006. Visual motion and the neural correlates of event perception. *Brain Research*, *1076*1, 150–162. https://doi.org/10.1016/j.brainres.2005.12.122

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., & Hasson, U. 2017. How We Transmit Memories to Other Brains: Constructing Shared Neural Representations Via Communication. *Cerebral Cortex*, *27*10, 4988–5000. https://doi.org/10.1093/cercor/bhx202

Zola-Morgan, S., Squire, L. R., & Amaralt, D. G. 1986. Human Amnesia and the Medial Temporal Region: Enduring Memory Impairment Following a Bilateral Lesion Limited to Field CA1 of the Hippocampus. *The Journal of Neuroscience*, *6*10, 18.

Zuo, X., Honey, C. J., Barense, M. D., Crombie, D., Norman, K. A., Hasson, U., & Chen, J. 2020. Temporal integration of narrative information in a hippocampal amnesic patient. *NeuroImage*, *213*, 116658. https://doi.org/10.1016/j.neuroimage.2020.116658

Zwaan, R. A., Langston, M. C., & Graesser, A. C. 1995. The Construction of Situation Models in Narrative Comprehension: An Event-Indexing Model. *Psychological Science*, *6*5, 292–297. https://doi.org/10.1111/j.1467-9280.1995.tb00513.x

Zwaan, R. A., & Radvansky, G. A. 1998. Situation Models in Language Comprehension and Memory. *Psychological Bulletin*, *123*2, 162–185. https://doi.org/10.1037/0033-2909.123.2.162