

Perception and encoding of narrative events during continuous speech listening in background noise

Ryan A. Panella^{a,b,*}, Alexander J. Barnett^{b,c}, Yulia Lamekina^a, Morgan D. Barense^{a,b}, Björn Herrmann^{a,b,*}

^a Rotman Research Institute, Baycrest Academy for Research and Education, North York, ON, Canada, M6A 2E1

^b Department of Psychology, University of Toronto, Toronto, ON, M5S 1A1, Canada

^c Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, Montreal, QC, H3A 2B4, Canada

ARTICLE INFO

Keywords:

Speech-in-noise
Event segmentation
Memory recall
Artificial intelligence
Narrative encoding

ABSTRACT

Everyday speech comprehension involves interpreting overarching themes as they unfold continuously overtime and structuring experiences for effective encoding and future recall. Characterizing downstream cognition, particularly how the brain perceptually organizes and encodes information, can help elucidate some of the mechanisms underlying listening effort. Recent advances in memory research highlight *event segmentation*, the process of identifying distinct events within dynamic environments, as a core mechanism to how we perceive, encode, and recall experiences. In the current study, we examine how background noise affects event segmentation during speech listening and its downstream effects on memory. Participants listened to and segmented narratives presented at varying signal-to-noise ratios (clear, +2 dB SNR, −4 dB SNR) and subsequently completed a free recall task. Increasing background noise reduced segmentation consistency and recall accuracy, indicating that challenging acoustics disrupt perceptual organization and encoding; however, listeners continued to identify meaningful event boundaries even when intelligibility declined by approximately 30%. Our analyses further suggest that segmentation was predominantly anticipatory, with listeners marking event boundaries toward the end of an event, suggesting a proactive updating of event models rather than a reaction to a prediction error. Additionally, segmentation-recall coupling was strongest under moderate noise, implying that moderate listening difficulty may enhance engagement. Ultimately, these findings demonstrate that while the adverse conditions impair detailed encoding, the cognitive mechanisms that structure experience remain robust, offering insight into how listening effort shapes perception and memory in complex, real-world speech.

1. Introduction

Speech comprehension is an intricate process that relies on the ability to decode, organize, and interpret complex auditory signals in real time. Auditory information unfolds continuously, requiring listeners to integrate acoustic features and contextual information to derive meaning. The temporal nature of auditory processing introduces unique challenges, particularly in environments where background noise or competing sounds are present. Everyday speech often occurs under such conditions (Hodgson et al., 2007; Mansour et al., 2021; Rusnock & Bush, 2012; Smeds et al., 2015), making comprehension effortful due to the additional cognitive resources needed to understand speech (Herrmann & Johnsrude, 2020a; Peelle, 2018; Pichora-Fuller et al., 2016). Previous works on speech comprehension have often

relied on the verbatim repetition of disconnected words or sentences that lack contextual cues and are not particularly engaging to listeners (Alhanbali et al., 2017; M.H. Davis & Johnsrude, 2003; Lunner & Sundewall-Thorén, 2007; Wendt et al., 2016; Zekveld et al., 2011). Although such designs are useful for capturing speech intelligibility, they offer limited insight into how speech is comprehended in everyday life amidst background noise. In natural settings, speech often resembles narratives where mental representations are rarely verbatim but rather reflect the gist across longer stretches of discourse. Indeed, listeners can remain absorbed and enjoy listening to continuous speech despite missing or mishearing a substantial proportion of words ($\geq 20\%$, Herrmann & Johnsrude, 2020b; Irsik et al., 2022a, 2022b; Yasmin et al., 2023). Understanding the challenges listeners – particularly older adults – face in everyday communication, requires novel approaches beyond

* Corresponding authors at: Rotman Research Institute, Baycrest Academy for Research and Education, 3560 Bathurst St., North York, ON, M6A 2E1, Canada.
E-mail addresses: ryan.panella@utoronto.ca (R.A. Panella), bherrmann@research.baycrest.org (B. Herrmann).

<https://doi.org/10.1016/j.actpsy.2026.107543>

Received 14 April 2026; Received in revised form 15 June 2026; Accepted 23 July 2026

Available online 30 July 2026

0001-6918/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

disconnected sentences and intelligibility measures. Instead, approaches are needed that capture comprehension and free recall of continuous speech that reflect natural conversational experience.

Insight into how we manage auditory complexity might come from theories of event perception and segmentation. Event segmentation theory (EST) proposes that individuals segment continuous experiences into meaningful *events* (Ezzyat & Davachi, 2011; Reynolds et al., 2007; Zacks et al., 2007; Zacks & Tversky, 2001). In cognitive psychology, an event can be described as an experience that occurs within the human scale that an observer perceives as having a beginning and an end (Barker, 1963; Barker & Wright, 1954; Dickman, 1963; Zacks et al., 2001). According to this framework, *event models* are constructed as working memory representations of the current experience, and through prior knowledge, are updated in response to prediction and prediction error, allowing the system to anticipate how an event likely unfolds (Kurby & Zacks, 2018; Zacks et al., 2007). Event models are constructed through the interaction between the sensory input and previous experiences, extrapolating on future information in real time; however, when perceptual details are no longer congruent with the current event model, this triggers the subsequent updating of the model in working memory, and the previous event is encoded in long-term memory (Boltz, 1992; Kurby & Zacks, 2008; Reynolds et al., 2007; Sargent et al., 2013; Zacks et al., 2007). Event boundaries mark transitions between episodes, scaffolding the organization of experience and supporting downstream memory encoding and recall (Kurby & Zacks, 2008; Lee & Chen, 2022; Swallow et al., 2009; Zacks et al., 2001). Speech comprehension requires listeners to engage in multiple layers of segmentation; however, event segmentation is distinct from lexical segmentation – the process of parsing continuous speech into individual words – which has been extensively studied in the auditory and linguistic sciences (Cutler & Butterfield, 1992; Cutler & Norris, 1988; Goh & Wallace, 2018). Whereas lexical segmentation operates at the level of individual words, event segmentation organizes extended experience into meaningful episodes, representing a higher-level cognitive process that is the focus of the current research.

Event segmentation has been extensively studied using both visual (e.g., silent films, written materials; E.E. Davis & Campbell, 2022; Newton et al., 1977; Newton & Engquist, 1976; Zacks et al., 2009) and audiovisual materials (e.g., films; Baldassano et al., 2017; Barnett et al., 2024; J. Chen et al., 2017; Lee & Chen, 2022; Speer & Zacks, 2005; Zacks et al., 2001), with comparatively fewer studies examining auditory-only materials (Burunat et al., 2024; Mei et al., 2024; Michelmann et al., 2025; Racciah et al., 2023). Across these paradigms, participants typically identify event boundaries during stimulus presentation by marking transition between events on paper or via button presses (Newton, 1973; Newton & Engquist, 1976; Zacks et al., 2009). Although event processing is inherently a subjective experience (Sargent et al., 2013), it has been shown that participants have the tendency to identify boundaries at similar locations, known as *normative boundaries* (Newton et al., 1977; Sasmita & Swallow, 2022; Zacks et al., 2006). Individuals whose segmentation patterns closely align with these normative boundaries appear to be attuned to meaningful narrative structure. Such alignment has been associated with stronger cognitive abilities (i.e., working memory capacity, perceptual speed, etc.) and enhanced memory of those events (Kurby & Zacks, 2011; Sargent et al., 2013; Swallow et al., 2009). The ability to segment in a manner that reflects the shared structure of an experience may serve as an index of how effectively information is being encoded and organized into long-term memory.

Although event segmentation has been well-characterized in visual and audiovisual contexts, much less is known about how individuals segment during naturalistic speech listening (Kumar et al., 2023; Michelmann et al., 2021; Zacks et al., 2009). Speech unfolds continuously, cannot be revisited, and background sound often requires the listener to recruit additional cognitive resources to maintain comprehension and track unfolding meaning (Baddeley & Hitch, 1974; Herrmann & Johnsrude, 2020a; Mattys et al., 2012; Pichora-Fuller et al.,

2016). The increased listening effort may have negative downstream consequences on working memory (Mattys et al., 2012; Pichora-Fuller et al., 2016; Rönnberg et al., 2013), suggesting a potential pathway through which acoustic challenges may influence event segmentation, by constraining the cognitive resources which are essential for maintaining and updating event models (Zacks et al., 2007; Zwaan et al., 1995; Zwaan & Radvansky, 1998). Because event boundaries are critical for structuring memory, segmentation impairments imposed by background noise may weaken the representations that support later recall – even when individual words are successfully understood.

Crucially, however, the type of recall most relevant in everyday life is not a verbatim reproduction of sentences, but rather the ability to reconstruct and communicate the gist of what was heard. In real-world conversations, successful comprehension may be judged by whether a listener can retell information, not whether they can repeat individual words exactly. Yet, the free recall of continuous speech, particularly in challenging listening conditions, has rarely been examined in the hearing sciences. In the current study, we examine how background noise impacts the perceptual organization of spoken narratives, with a focus on the relationship between event segmentation and memory. Based on the event segmentation theory, we expect that more effective segmentation will be associated with improved recall of narratives content. We further predict that challenging listening conditions may disrupt the relationship by reducing comprehension at encoding and taxing the cognitive resources required to maintain and update event models. Given the novelty of examining event segmentation during naturalistic speech listening, several aspects of the present work are intentionally exploratory. Specifically, we examine the temporal dynamics of segmentation responses, including when participants tended to identify event boundaries relative to a meaningful narrative transition (i.e., normative boundaries), to better characterize how listeners track unfolding narratives under varying degrees of acoustic challenge. These analyses are intended to generate hypotheses about the mechanisms linking listening effort, event model maintenance, and memory encoding. To support this investigation, we capitalize on recent advances and validations in the use of large language models (LLMs) to support and automate analyses that previously required human scoring, including the identification of normative event boundaries and the scoring of free recall (Herrmann, 2025a; Panella et al., 2025). This promotes a scalable approach to studying comprehension and memory in naturalistic paradigms. Ultimately, such insights are essential for developing a comprehensive understanding of how humans successfully navigate the complexities of everyday communication.

2. Methods

2.1. Participants

Thirty-six younger adults participated in the current study ($M_{\text{age}} = 23.6$, range: 18–34 years, $N_{\text{female}} = 23$, $N_{\text{male}} = 13$). The sample size was determined based on prior event segmentation research, which has demonstrated robust stability in segmentation patterns with similar or smaller sample sizes (Sasmita & Swallow, 2022). Six additional participants were excluded: three due to poor performance on the intelligibility task potentially indicating undiagnosed hearing impairments (proportion words correct < 0.8 , normal-hearing younger adults are expected to perform near ceiling under clear listening conditions; Irsik et al., 2022a; Wilson et al., 2012) and three due to failure to register any event boundary responses during the task. Participants were either native English speakers or highly proficient English speakers, having learned English before age five. All participants reported having no known neurological diseases. Participants provided written consent prior to the study and received compensation at \$10 per 30-min of participation or through course credit at the University of Toronto. The study was conducted in accordance with the Declaration of Helsinki (Version 2004) and the Canadian Tri-Council Policy Statement on

Ethical Conduct for Research Involving Humans (TCPS2–2014), and was approved by the Research Ethics Board of the Rotman Research Institute at Baycrest Academy for Research and Education (REB #23–11). This study was not preregistered.

2.2. Stimuli

We used narratives extracted from the official audiobook of Trevor Noah's memoir *Born a Crime* (Noah, 2016). The book highlights his experiences growing up during the era of Apartheid in South Africa. Each chapter provides a cohesive narrative intended to create an absorbing and enjoyable reading experience for the participant. We used the following three chapters: Run!, Go Hitler, and My Mother's Life. The audiobook chapters were truncated such that each narrative was approximately 10 min (~1500 words) while maintaining narrative closure. Each chapter served as an independent, self-contained narrative with a distinct setting, characters, and progression, such that comprehension of one chapter did not depend on familiarity with the others. Additionally, the chapter Robert, was used as a practice narrative to introduce the experimental procedure (~3 min; Fig. 1). Prior to stimulus selection, familiarity with the source material was informally assessed among potential participants and the book was found to be not widely known within our participant pool. Participants were not informed that the narratives were drawn from a published memoir prior to completing the task.

To investigate the impact of speech masking – a proxy for listening effort – on event segmentation patterns and recall accuracy, participants were presented with three narratives under distinct listening conditions. One of the narratives was presented under clear conditions (i.e., without background noise) while the remaining two stories were overlaid with a 12-talker babble at a signal-to-noise ratio (SNR) of +2 dB and – 4 dB, respectively (Bilger et al., 1984; Wilson et al., 2012). Twelve-talker babble simulates a crowded environment, but does not allow for the identification of individual words (Bilger et al., 1984; Mattys et al., 2012).

2.3. Experimental procedure

2.3.1. Event boundary identification and recall

The experimental procedure involved participants listening to a narrative and identifying event boundaries in real time through button presses and subsequently performing a free recall of the narrative. Instructions were presented through three modalities to ensure clarity: printed handouts, verbal guidance, and visual cues on a computer screen using PsychoPy 2023.2.3 (Peirce et al., 2019). Auditory stimuli were presented at a comfortable listening level that was fixed across participants (~70 dB SPL). Participants were informed that the practice block would involve listening to the same narrative two times under clear speech conditions and that their primary task involved identifying events within the narrative. This was designed to familiarize them with

the task demands prior to introducing degraded listening conditions. Specifically, an *event* was defined as a discrete piece of information which could be described as having a clear beginning and an end. To demarcate events, participants were instructed to press the spacebar whenever they perceived one event concluding and another beginning. Importantly, participants were told that there were no objectively correct answers and that responses were subjective, allowing for individual variation. Across the two narrative listens, participants were asked to identify both small and large event units (Kurby & Zacks, 2011; Swallow et al., 2009, 2018; Zacks et al., 2001). In the first listen, they focused on the smallest natural and meaningful event units. In the second reading, they shifted their attention to larger event units. This dual approach aimed to provide participants with a nuanced understanding of narrative levels (Kurby & Zacks, 2011; Panella et al., 2025; Swallow et al., 2009; Zacks et al., 2001) as the main procedure instructed participants to focus solely on the large event units (Panella et al., 2025).

After completing the segmentation task, participants provided a free recall of the practice narrative. They were instructed to speak into a microphone (Shure SM7B; Steinberg UR22C external sound card) and provide as much detail as possible, even if they believed some details were insignificant to the narrative progression. The free recall for the practice narrative was performed twice. Participants first recalled the narrative and once they indicated they were complete, a second screen prompted them to add any final details they might have missed in their initial recall. This approach during practice helped participants to provide additional information they would have otherwise not reported and recognize the depth of recall we were expecting (Karpicke & Roediger, 2008; Panella et al., 2025; Ratcliff, 1978; Roediger & Butler, 2011; Roediger & Karpicke, 2006). They were instructed to provide a full recall of all details remembered during the main experimental procedures.

Once participants demonstrated an understanding of the tasks, they proceeded to the main experiment. Participants listened to the narratives and identified event boundaries for each of the three narratives, one narrative at the time (i.e., in different blocks with appropriate breaks in between). Each of the narratives were presented in one of the three listening conditions (i.e., clear, +2 dB SNR, – 4 dB SNR). The order and noise conditions in which the three narratives were presented was counterbalanced across participants. Unlike the practice task, participants were instructed to focus only on the large event units. The free-recall phase immediately followed each narrative segmentation phase. They were encouraged to provide exhaustive detail, but unlike the practice task, they were only provided one opportunity for recall.

2.3.2. Speech intelligibility assessment

After completing the main experiment, participants performed a traditional speech intelligibility task to assess their ability to identify spoken words under varying noise conditions (i.e., word report). This task served both to characterize participants' speech-in-noise performance and as a quality-control measure for inclusion in subsequent analyses. Thirty sentences were extracted from the audiobook chapter *Loopholes* by Trevor Noah (Noah, 2016). Sentences were randomly assigned to one of the three noise conditions used in the main experiment (i.e., clear, +2 dB SNR, – 4 dB SNR), with ten sentences per condition. The assignment of sentence to noise conditions were randomized across participants. Each sentence was presented once, and the order of presentation was randomized across participants. Participants were instructed to listen carefully and type exactly what they heard into a textbox. Typed responses are standard in the speech intelligibility literature (Herrmann, 2025b; Irsik et al., 2022a; Pandey & Herrmann, 2025) and were selected for their compatibility with automated scoring. No feedback was provided. Accuracy was later assessed by comparing typed responses to the original transcripts using exact string matching, typical of previous work in the auditory sciences (Allison, 2020; Allison et al., 2025; Flanagan, 1972; Herrmann, 2025b; Hustad et al., 2015; Irsik et al., 2022a; Kent et al., 1989).



Fig. 1. Experimental Design. Participants initially segmented and recalled a short narrative to ensure they understood instructions. Subsequently, they listened to three narratives, identifying the largest event units throughout, and immediately freely recalled the narratives by speaking into the microphone.

2.4. Analysis

Analyses are organized to reflect the primary research questions regarding the effects of listening effort on event segmentation and downstream consequences for narrative recall. While some analyses are motivated by prior work and theoretical predication, others are more exploratory in nature, including analyses characterizing the temporal dynamics of boundary responses and investigating potential mechanisms linking segmentation performance and memory. Specifically, we first report the elementary effects of background noise on speech intelligibility, followed by analyses of event segmentation behaviour, including event boundary counts and segmentation agreement. We subsequently report analyses of narrative recall, including recall performance and temporal organization. Additional novel and exploratory analyses examining the precise timing of boundary responses relative to normative event boundaries, boundary salience, the relationship between segmentation and recall, and event centrality are reported. Specific analytical methods are described in their respective sections below.

2.4.1. General statistical methodology

Statistical analyses were performed using R 4.4.0 (R Core Team, 2024). Linear and generalized linear mixed effects models were conducted using *lme4* (Bates et al., 2015). Linear models were interpreted using ANOVA tables through *lmerTest* (Kuznetsova et al., 2017). Categorical variables were entered using the default factor coding, with omnibus F-test used to assess their significance as a predictor. Wald χ^2 tests were conducted for the generalized linear mixed-effects models using the *car* package (Fox & Weisberg, 2019) and effect sizes were generated using the *effectsize* package (Ben-Shachar et al., 2020). Post hoc analyses were conducted using *emmeans* (Lenth, 2024) with Tukey adjusted *p*-values (denoted as *p_T*) and Kenward-Roger approximated degrees of freedom (Kenward & Roger, 1997). Exact *p*-values are reported when possible and approximated to common conventions when values are subject to computational underflow (i.e., *p* < .0001).

The random effects structure in the linear mixed effects models was tailored to the design of each analysis. For within-subject analyses, we included random intercepts for subject and narrative to account for variability at both levels. This is noted and justified in the corresponding methods sections. We adopted a maximal random effects approach, including random slopes where justified by the design and theoretical motivation (Barr, 2013; Barr et al., 2013). Models that encountered singular fits were subsequently simplified to ensure model stability.

The analyses reported in the current study address multiple related research questions across distinct dependent variables, rather than testing a single a priori hypothesis. While some analyses are motivated by prior work and theoretical predictions, others are more exploratory in nature. Specific analytical methods are outlined in their respective sections below.

2.4.2. Analysis of speech intelligibility data

Speech intelligibility responses were assessed by calculating the proportion of correctly identified words (Borrie et al., 2019; Herrmann, 2025b; Pandey & Herrmann, 2025). Participant response transcripts were scored using *Autoscore*, an open-source R package designed for automated scoring of listeners' transcripts (Borrie et al., 2019), using all default parameters. *Autoscore* compares participant transcripts to target sentences, applying customizable rules to account for common misspellings, grammatical variations, and other acceptable deviations, mitigating potential disadvantages for participants with spelling difficulties. This analysis yielded a proportion of words correct score for each sentence. A linear-mixed effects model was fit using the proportion of words correct as the dependent measure, noise conditions as the within-subject fixed effect, and a random intercept was included for participant.

2.4.3. Analysis of event segmentation data

2.4.3.1. Number of events boundaries. For each participant, the number of identified boundaries were counted for each narrative. To account for differences in narrative length, the number of identified boundaries was divided by the duration of the corresponding narrative and scaled to a per-60 s basis to standardize the comparisons. A linear mixed effects model was calculated using the scaled number of events as the dependent variable and noise condition (i.e., clear, +2 dB SNR, -4 dB SNR) as the within-subjects factor. Random intercepts were included for participant and narrative.

2.4.3.2. Segmentation agreement index. To analyze the degree to which participants agreed in their placement of event boundaries, we calculated a segmentation agreement index (Kurby & Zacks, 2011; Sargent et al., 2013; Sasmita & Swallow, 2022; Swallow et al., 2018; Zacks et al., 2006). For each participant, a binary time series was generated over the duration of each narrative at a 0.01-s resolution, marking a one at each timepoint where an event boundary was identified and a zero otherwise (Sasmita & Swallow, 2022; Zacks et al., 2001, 2006). These timeseries were then convolved using a Gaussian kernel with a bandwidth of five seconds via the *ksmooth* function from the *stats* package in R (R Core Team, 2024) to accommodate temporal imprecision in responses. This resulted in a timeseries for each narrative/noise condition per participant. We employed a leave-one-out approach, calculating the Spearman correlation between the participant's timeseries and the averaged timeseries across all other participants listening to the same story under the same noise conditions (Kurby & Zacks, 2011; Sargent et al., 2013; Sasmita & Swallow, 2022; Swallow et al., 2018; Zacks et al., 2006). A linear mixed effects model was fit using the agreement index as the dependent variable and noise condition as the fixed effect, with participant and narrative included as random intercepts.

2.4.3.3. LLM derived normative event boundaries. Although event segmentation is a subjective experience, prior research has shown that individuals tend to identify event boundaries at consistent points in a narrative — locations referred to as *normative boundaries* — suggesting the presence of shared contextual cues that mark meaningful transitions (Newtonson & Engquist, 1976; Sasmita & Swallow, 2022; Zacks et al., 2006).

Although participant-averaged boundary responses are a common normative reference in event segmentation research (Barnett et al., 2024; Sasmita & Swallow, 2022), deriving normative event boundaries from the same participants whose responses are being analyzed would introduce circularity in analyses. LLM-derived event boundaries provide an independent reference that avoid this concern, and recent work suggests that such boundaries align with human judgments even more consistently than humans align with each other (Michelmann et al., 2025; Panella et al., 2025). Accordingly, we leveraged OpenAI's GPT-4 to identify normative event boundaries for precise time locked analyses (Panella, 2025). Specifically, verbatim transcripts of each narrative were stripped of formatting, such as paragraph breaks and punctuation cues, to minimize bias in segmentation, and were then separately submitted to GPT-4 through the application processing interface (API) using Python 3.11.5 (van Rossum & Drake, 2010). Using a zero-shot prompt (B. Chen et al., 2025) with a temperature parameter of 0, GPT-4 was instructed to reproduce the narrative word-for-word and segment the narrative into large events (see Michelmann et al., 2025; Panella et al., 2025). Each narrative was independently submitted to GPT-4 twenty times to account for potential variability in output. From these twenty iterations, we computed the number of times a boundary was placed at each point in the text. Normative event boundaries were defined as the *n* most frequently identified boundaries, where *n* was equal to the mean number of boundaries produced across the twenty iterations for that story (Panella et al., 2025; Sasmita & Swallow, 2022).

All identified normative boundaries were located at the end of sentences.

Critically, the LLM procedure was used solely to identify which sentence transition constituted normative event boundaries. The precise temporal reference points used in subsequent analyses – event offset and event onset – were defined acoustically via forced alignment and therefore reflect objective acoustic landmarks independent of any LLM judgement. To obtain the timing of these normative event boundaries in the auditory narratives, we used the online implementation of CLARIN's forced alignment software (Kisler et al., 2017). The aligner takes a verbatim transcript and audio file as input and returns onset times for each word in the spoken narrative. Output files were manually reviewed, and any misalignments were corrected to ensure accuracy. The offset time of the last word preceding and onset time of the first word following each LLM-identified boundary was recorded. In addition to these boundary timepoints, we computed event centres – the temporal midpoint between two consecutive boundaries, rounded to the nearest sentence end.

2.4.3.4. Time locked event boundary responses. To examine how background noise affects participants' sensitivity to meaningful narrative transitions, we conducted a series of exploratory time locked event boundary analyses. These analyses go beyond traditional segmentation metrics by examining not only whether participants identified normative event boundaries, but precisely when they did so relative to the acoustic structure of the narrative. These analyses addressed three questions: (1) whether responses clustered around specific, contextually driven transition points in the narrative, (2) whether background noise modulated the magnitude of these responses, and (3) whether background noise influenced the timing of boundary detection.

To quantify participants' alignment with normative boundaries, we used the LLM identified event boundaries and corresponding event centres to compute the proportion of boundary responses that occurred temporally near each normative event boundary and event centre. Specifically, for each participant, we implemented a sliding window approach across a time range from -10 to $+10$ s relative to each LLM-identified normative boundary and event centre. We calculated the number of button presses falling within each 3-s window, shifting the window in 0.05-s steps to generate a continuous time-resolved profile. For each participant, narrative, and noise condition, the number of presses within each window was normalized by the total number of normative boundaries, yielding a proportion that reflects the probability of identifying an event boundary over time.

Because spoken speech comprises brief pauses and intonational breaks between speech segments (Kahng, 2018; Liu et al., 2022; Owoicho et al., 2024; Rochester, 1973; Zellner, 1994), the offset of one segment (e.g., final word in the sentence) and the onset of the subsequent segment (e.g., first word in the next sentence) are temporally distinct timepoints, separated by a natural inter-sentence pause ($M_{\text{duration}} = 1.06$ s, $SD = 0.33$ s, range: 0.54–2.07 s). Specifically, our normative event boundaries were defined as the interval between successive events. This temporal structure, therefore, applies to event boundaries and event centres, which coincided with transitions between successive sentences. This feature of spoken speech means that participant responses can occupy a different temporal position depending on the reference point. For instance, a response occurring within the inter-sentence pause would be classified as post-boundary when responses are time-locked to event offset, but pre-boundary when time-locked to event onset. The event segmentation theory suggests that segmentation is triggered by a mismatch between incoming information and the current event model (Nguyen et al., 2024; Nolden et al., 2024; Reynolds et al., 2007; Zacks et al., 2007, 2011). Separately aligning boundary responses to these timepoints could provide an exploratory characterization of whether participants tended to identify event boundaries based on the anticipation of an impending narrative transition (anticipatory

segmentation; responses aligned to event offsets) or the failure to integrate incoming information (reactive segmentation; responses aligned to event onsets). Although these analyses cannot directly distinguish predictive from reactive mechanisms, they may provide insight into the temporal dynamics of boundary detection. If the peak boundary-identification time occurs around or before the onset of a new event, this could suggest that participants anticipate an upcoming transition rather than reacting to new information in the subsequent event.

In the first analysis, to determine whether responses preferentially tracked meaningful narrative transitions, we collapsed across noise conditions and compared participant responses time locked to normative event boundaries versus event centres. This allowed us to assess whether segmentation responses were selectively elevated around true narrative transitions, as opposed to within-event timepoints, regardless of listening difficulty. This analysis also provides confirmation that LLMs can successfully identify normative boundaries for auditory narrative listening (for narrative reading see Panella et al., 2025). For normative event boundaries, we calculated the average proportion of identified boundaries in predefined pre- and post-boundary windows. In the event onset condition the pre-boundary window spanned -3 to -1 s, and the post-boundary window spanned 0 to 2 s. In the event offset condition, the pre- and post-boundary windows spanned -2 to 0 s and 1 to 3 s, respectively. For event centres, we used the same two-second windows for each condition, but rather time locked to the sentence transition approximately at the temporal midpoint between successive event boundaries. The event centre condition matched the event boundary condition in window duration and relative timing, differing in only the temporal reference (within-event midpoint vs. narrative transition). Linear mixed effects models were fit to assess the proportion of identified boundaries, separately for the event offset and onset conditions. Fixed effects included the temporal reference (i.e., event boundary, event centre), relative time window (i.e., pre-boundary, post-boundary), and their interaction. Random intercepts were included for participant and narrative.

Having established that participant responses were clustered around event boundaries, we conducted a second analysis to evaluate how background noise modulated the magnitude of segmentation patterns. Unlike the previous analysis, we solely used the event offset and extracted the average proportion of identified boundaries from the predefined pre- and post-boundary windows as our analyses suggest that participants may anticipate event endings rather than reacting to the beginning of a new event: -2 to 0 s and 1 to 3 s, respectively. To account for baseline variability, we subtracted the average pre-boundary amplitude from the post-boundary amplitude for each participant and noise condition. By correcting for pre-boundary responses, this measure specifically isolates boundary-evoked segmentation responses and reduces the influence of unrelated responses. The primary variable of interest was noise condition (i.e., clear, $+2$ dB SNR, -4 dB SNR). A linear mixed effects model was used to predict the proportion of identified event boundaries as a function of noise condition with random intercepts included for participant and narrative.

Finally, in addition to the magnitude of the proportion of identified boundaries, we also examined response latency to assess whether background noise shifted the timing of boundary detection. To estimate segmentation latency, we calculated the centre of mass of each participant's response distribution following each event boundary (Clayson et al., 2013; Dien et al., 2004). Latency was calculated using responses in the 0 to 6 s window following the offset of the event. We specifically computed the weighted average of timepoints, where each timepoint was weighted by the proportion of boundary responses it contained (Clayson et al., 2013; Dien et al., 2004). This approach provides a robust estimate of when boundary responses were concentrated, accounting for both timing and magnitude. Narratives for which participants did not indicate any boundaries within the post-boundary window were excluded from the latency analysis. This occurred in 6% of blocks across all participants but more frequently in the -4 dB SNR condition (13%)

than clear (0%) and +2 dB SNR (5%). A linear mixed effects model was used to assess whether response latency varied across noise conditions, with participant and narrative included as random intercepts.

2.4.3.5. Normative boundary salience index. As an additional exploratory analysis, we examined whether the salience of normative event boundaries predicted participants' likelihood of detecting that boundary – and whether this relationship was modulated by listening difficulty. This measure quantified the extent to which some event boundaries were more likely to be identified than others, based on a normative reference. For each normative boundary identified by the LLM, boundary salience was operationalized as the proportion of LLM iterations (out of 20) that included a boundary at that location. For each participant and narrative, we then determined whether a button press occurred within 6 s of the end of the offset of the previous event, defining such cases as successful event boundary detection. We computed a logistic mixed effects regression model which was used to predict boundary identification based on the salience of the boundary (i.e., LLM agreement proportion), noise condition (i.e., clear, +2 dB SNR, –4 dB SNR), and their interaction. Random intercepts were included for participant and narrative, with a by-participant random slope of normative boundary salience. This analysis allowed us to quantify how the salience of normative event boundaries and listening conditions jointly modulates the likelihood of listeners detecting event boundaries.

2.4.4. Analysis of recall data

2.4.4.1. Transcription and segmentation of recorded recall. Spoken recall data was used for the following analyses. Recall audio was initially transcribed through Otter.ai (Corrente & Bourgeault, 2022) and a manual transcriber verified and cleaned all transcripts. To directly relate event segmentation to recall, we employed a validated automated event recall scoring approach (see Panella et al., 2025). For each recall transcript, normative boundaries were identified using the same automated event segmentation approach with OpenAI's GPT-4 model as described above for narrative text. These boundaries were used to break the recall transcript into segments, resulting in recall events that were used for recall scoring described below.

2.4.4.2. Semantic similarity matrix. The current approach utilized semantic embeddings, specifically Google's Universal Sentence Encoder (USE, Cer et al., 2018) to assess the similarity between narrative and recall events. The USE encodes text into high-dimensional vectors, where vectors for semantically similar texts show higher correlations than vectors for texts that are less semantically similar. The same LLM derived normative event boundaries for each narrative previously generated for the event segmentation analyses were utilized in the recall analyses.

For each participant and narrative, a semantic similarity (i.e., correlation) matrix was generated by calculating the Spearman correlations between the text-embedding for each narrative segment and the text-embedding for each recall segment (Herrmann, 2025a; Heusser et al., 2021; Panella et al., 2025; Raccach et al., 2024; Shen et al., 2023). This resulted in a rectangular matrix, as the number of narrative and recall events often differed. To allow for consistent comparisons across participants and support downstream analyses, the matrix was transformed into a square format of size $n \times n$, where n is the number of narrative events (Herrmann, 2025a; Panella et al., 2025) using the `resize` function from the *Scikit-image* package (Van Der Walt et al., 2014) in Python. This function rescales the matrix to the target dimensions using bilinear interpolation, which continuously redistributes similarity scores across the new matrix dimensions regardless of whether the recall contained fewer or more events than the narrative. The resulting square matrix (i.e., narrative $\times$ recall matrix) reflects the alignment between narrative and recall content across corresponding event segments. The semantic

similarity matrix serves as the basis for various metrics described in more detail below.

2.4.4.3. Event recall scores. For each narrative event in the narrative $\times$ recall matrix (i.e., for each row), we identified the maximum correlation value, representing the best matching recall event and subsequently serving as the recall score for that narrative event (Herrmann, 2025a; Panella et al., 2025). This approach has recently been validated, showing correspondence with recall scoring conducted with humans (Panella et al., 2025). To determine whether the event recall scores exceeded chance level, we first computed a baseline condition by assessing recall transcripts against the two non-corresponding narratives. For each comparison, the recall matrix was reshaped to a square matrix relative to the number of events in the corresponding narrative. As described above, the maximum semantic similarity score per narrative event was extracted and averaged across the two non-corresponding narratives to obtain a chance-level estimate. We conducted a linear mixed effects analysis invariant of noise condition with score type (i.e., observed recall score, chance recall score) as the fixed effect. Random intercepts were included for participant and narrative, along with a by-participant slope for score type. A subsequent linear mixed effects model was fit to determine the effects of background noise on recall performance, using only the observed event recall scores as the dependent measure (i.e., chance level data were not included to avoid model complexity), modeling noise condition as the main predictor. Random intercepts were included for subject and narrative, and by-participant slope for noise condition.

2.4.4.4. Temporal order. Previous research suggests that individuals have the tendency to recall information in the order in which it was experienced (J. Chen et al., 2017; Herrmann, 2025a; Heusser et al., 2021; Howard & Kahana, 2002; Zhang et al., 2023). The diagonal elements from the semantic similarity matrix reflect the recall scores for the narrative events recalled in the correct temporal order. To determine whether participants' recall followed this forward trajectory, we extracted the diagonal elements along with their corresponding off-diagonal counterparts – specifically, the reverse-diagonal values representing the same event pairs recalled in opposite order. A linear mixed effects model was fit to predict the similarity scores as a function of temporal direction (i.e., forward, reverse), noise condition, and their interaction, along with random intercepts for subject and narrative.

Subsequently, we quantified the temporal divergence of recall to assess the degree to which participants diverged from recalling the narrative in perfect temporal order (Herrmann, 2025a; Heusser et al., 2021). We computed the root mean square error (RMSE) between the index of the event recall score (i.e., maximum correlation value for each narrative event – row in the narrative $\times$ recall matrix) and corresponding position along the diagonal of the similarity matrix, with higher values indicating greater out-of-order recall. Temporal divergence was analyzed as a function of noise condition using a linear mixed effects model, with random intercepts for participant and narrative.

2.4.4.5. Segmentation ability on recall. We estimated the unique contribution of segmentation ability on recall by conducting separate analyses with the agreement index and post-boundary amplitudes from the event offset analysis. Because segmentation covaried with noise condition, we first orthogonalized it by modeling segmentation as a function of noise condition through a linear mixed effects model including random intercepts for participant and narrative (García et al., 2020; Geldhof et al., 2013; Perin et al., 2014). The resulting conditional residuals were extracted and used as segmentation regressors. Subsequently, a linear mixed effects model was fit to predict event recall scores as a function of noise condition, the resulting segmentation residuals, and their interaction. Random intercepts were similarly included for participant and narrative.

2.4.4.6. Event centrality. Recent research suggests that events that are more central or important to a narrative may be recalled more effectively (Lee & Chen, 2022). Given the role of narrative structure in supporting recall, we aimed to examine whether central events were more resilient to the effects of listening effort. To obtain an event centrality score for each narrative, a self-similarity matrix was calculated for each narrative using the event embedding vectors. Each value in the resulting narrative $\times$ narrative matrix represents the similarity between two events, providing a comprehensive view of the relationships among all events in the narrative. For each narrative event, we calculated its centrality score by averaging the correlation score across its corresponding row in the self-similarity matrix. The average correlation score was deemed the centrality score, reflecting the event's overall connectedness and importance within the narrative. Higher centrality scores indicate events that are more semantically connected to other events, suggesting their pivotal role in the narrative's coherence and structure. A linear mixed effects model was fit to determine whether centrality scores can predict recall performance. Specifically, the model included centrality scores, noise condition (i.e., clear, +2 dB SNR, -4 dB SNR), and their interaction as fixed effects. Random intercepts were included for participant and narrative, along with a random slope for noise condition to account for participant-specific slopes over centrality.

3. Results

3.1. Speech intelligibility

Intelligibility decreased as background noise increased, as expected (smallest $t_{70} = 3.59$, $p_T = 1.75 \times 10^{-3}$; main effect of noise condition: $F_{2,70} = 238.05$, $p = 5.94 \times 10^{-32}$, $\eta_p^2 = 0.86$, 95% CI [0.82, 0.91]; Fig. 2A). Three participants were excluded from this and all subsequent analyses based on unexpectedly low intelligibility in the clear speech condition (proportion words correct <0.80). Clear speech should be highly intelligible for normal-hearing younger adults (Irsik et al., 2022a; Wilson et al., 2012), and we thus assume the three participants either had some undiagnosed hearing impairments or were unmotivated to perform the task.

3.2. Event segmentation

3.2.1. Number of event boundaries

A linear mixed effects model for the number of identified events

indicated a significant effect of noise condition ($F_{2,68.28} = 5.95$, $p = 4.14 \times 10^{-3}$, $\eta_p^2 = 0.15$, 95% CI [0.02, 0.30]; Fig. 2C). The -4 dB SNR condition resulted in the least number of events in comparison to both the +2 dB SNR and clear speech conditions (smallest $t_{68.5} = 2.94$, $p_T = 1.24 \times 10^{-2}$), but there were no significant differences between the latter conditions ($t_{68.3} = 0.06$, $p_T = 0.998$). This indicates that an increase in background noise reduces the average number of events identified per minute, but only in the hardest condition.

3.2.2. Segmentation agreement index

The agreement index measures how well a participant's event boundary responses aligned with the rest of the group. Fig. 2B illustrates this visually, showing the locations within the narrative where participants are most likely to identify an event boundary for each noise condition. This measure was greater for the +2 dB SNR and clear speech conditions in comparison to the -4 dB SNR condition (smallest $t_{68.5} = 4.93$, $p_T = 1.61 \times 10^{-5}$; main effect of noise condition: $F_{2,68.27} = 22.42$, $p = 3.28 \times 10^{-8}$, $\eta_p^2 = 0.40$, 95% CI [0.21, 0.54]; Fig. 2D). There were no differences in the agreement index between speech masked at a +2 dB SNR and clear speech ($t_{68.3} = 1.39$, $p_T = 0.352$). This suggests that participants are equivalently consistent in their response locations in easy and moderate listening conditions.

3.2.3. Time locked event boundary analyses

3.2.3.1. Temporal position of event boundary responses. We conducted a time locked analysis around normative event boundaries and event centres using a sliding window approach, calculating the proportion of boundary responses centred at both the event onset (i.e., reaction of the beginning of a new event) and offset (i.e., anticipation of a narrative transition). When time locked to event onset, participants showed a clear increase in the proportion of boundary responses in the post-onset window (0 to 2 s) relative to the pre-boundary window (-3 to -1 s; main effect of relative time window: $F_{1,380.93} = 36.68$, $p = 3.34 \times 10^{-9}$, $\eta_p^2 = 0.09$, 95% CI [0.04, 0.15]; Fig. 3A). This effect was stronger for event boundaries than for event centres, as indicated by a main effect of temporal reference ($F_{1,381.56} = 40.44$, $p = 5.80 \times 10^{-10}$, $\eta_p^2 = 0.10$, 95% CI [0.05, 0.16]) and a significant temporal reference $\times$ relative time window interaction ($F_{1,380.93} = 29.95$, $p = 8.05 \times 10^{-8}$, $\eta_p^2 = 0.07$, 95% CI [0.03, 0.13]). Post hoc comparisons confirmed that response proportions significantly increased from pre- to post-boundary windows for event boundaries ($t_{381} = 8.17$, $p_T < 0.0001$), but not for event centres

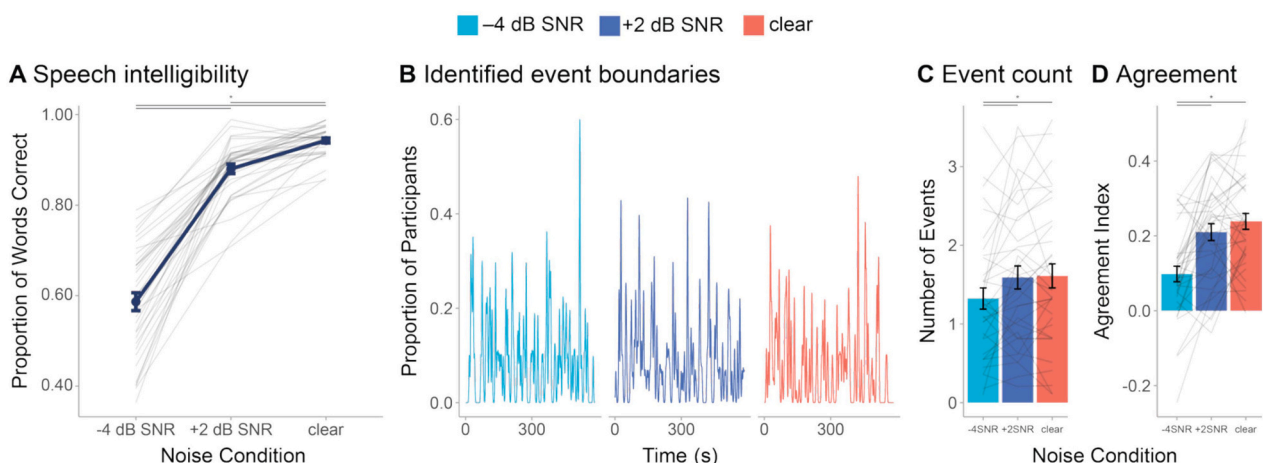


Fig. 2. Speech Intelligibility & Segmentation Results. A. Averaged participant intelligibility scored as a proportion of correctly reported words. Participants whose clear speech intelligibility was below 0.8, were excluded from subsequent analyses. Error bars represent ± 1 standard error of the mean (SEM) across participants ($n = 36$). B. Location of identified event boundaries in narrative one smoothed using a Gaussian filter with a bandwidth of five seconds. Each pane shows the proportion of participants that identified event boundaries at each second in the narrative. C. Number of identified event boundaries. Bars show the mean across participants per minute. D. Event segmentation agreement index. Error bars represent ± 1 SEM across participants ($n = 36$).

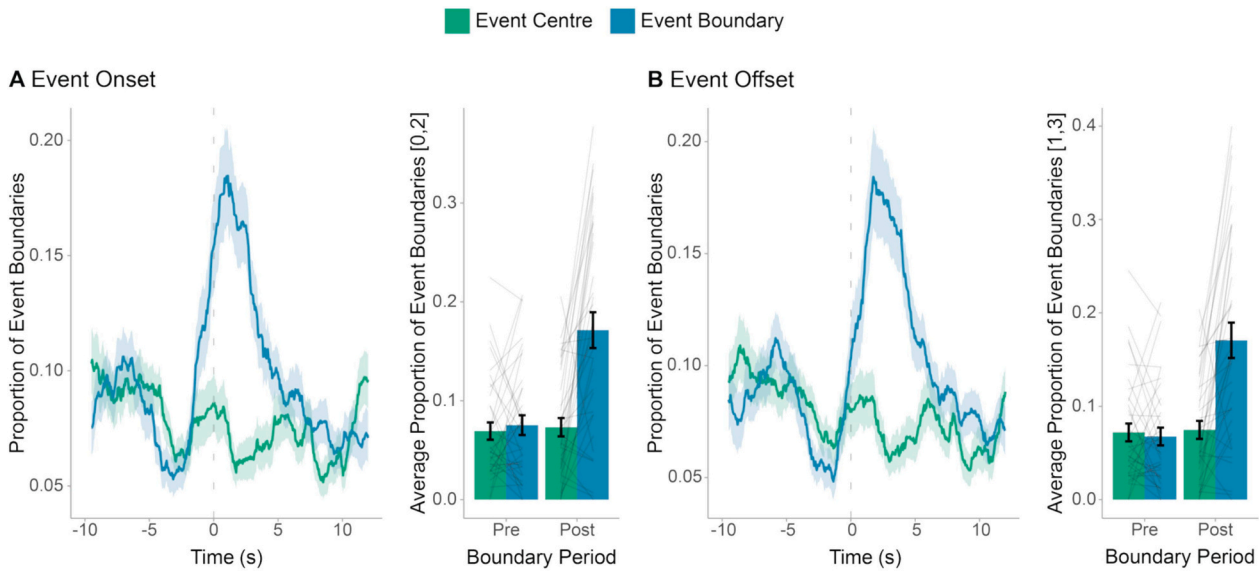


Fig. 3. Time Locked Event Boundary Responses. **A.** Proportion of identified event boundaries using a 3-s sliding window time locked to event onset. Bar graphs show the average proportion of events identified in pre-boundary (−3 to −1 s) and post-boundary (0 to 2 s) windows. **B.** Proportion of identified event boundaries using a 3-s sliding window time locked to event offset. Bar graphs show the average proportion of events identified in pre-boundary (−2 to 0 s) and post-boundary (1 to 3 s) windows. Shaded error ribbons and error bars represent ± 1 SEM across participants ($n = 36$).

($t_{381} = 0.412$, $p_T = 0.976$).

A similar pattern was observed when data were time locked to event offsets. Participants showed a significant increase in the proportion of boundary responses in the post-boundary window (1 to 3 s) compared to the pre-boundary window (−2 to 0 s; main effect of relative time window: $F_{1,380.97} = 39.74$, $p = 8.03 \times 10^{-10}$, $\eta_p^2 = 0.09$, 95% CI [0.05, 0.15]; Fig. 3B). This effect was again specific to event boundaries, as indicated by a main effect of temporal reference ($F_{1,381.58} = 30.10$, $p = 7.51 \times 10^{-8}$, $\eta_p^2 = 0.07$, 95% CI [0.03, 0.13]) and a significant temporal reference $\times$ relative time window interaction ($F_{1,380.97} = 34.28$, $p = 1.03 \times 10^{-8}$, $\eta_p^2 = 0.08$, 95% CI [0.04, 0.14]). Post hoc comparisons showed that

responses significantly increased from pre- to post-boundary windows for event boundaries ($t_{381} = 8.62$, $p_T < 0.0001$), but not for event centres ($t_{381} = 0.32$, $p_T = 0.989$).

These results show that participants are sensitive to the timing of meaningful narrative transitions identified by an LLM, confirming previous work for written narratives (Panela et al., 2025). Moreover, the proportion of button presses for data time locked to event onsets peaked around one second after event onset, whereas the proportion for data time locked to event offsets peaked at two seconds (Fig. 3). Little information is available within the one second region post-boundary that would enable a listener to determine a new event has started, suggesting

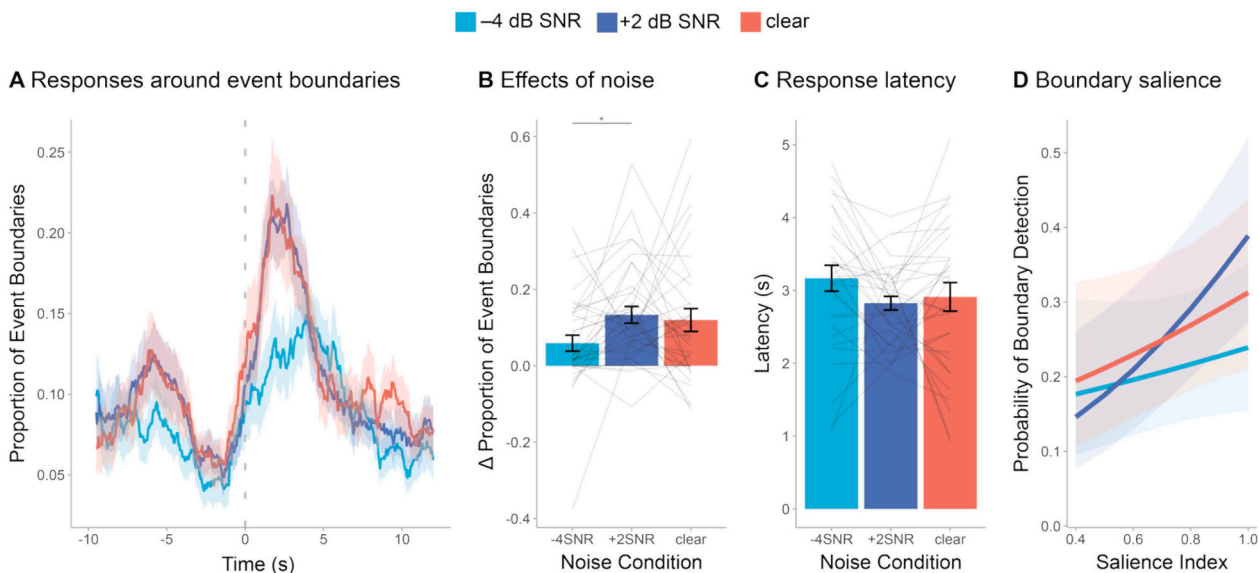


Fig. 4. Noise-Dependent Dynamics of Event Boundary Responses. **A.** Proportion of identified event boundaries time locked to event offset separated by noise condition. Shaded error ribbons represent ± 1 SEM across participants. **B.** Average change in the proportion of events identified from pre-boundary (−2 to 0 s) to post-boundary (1 to 3 s) windows. Error bars represent ± 1 SEM across participants. **C.** Weighted average latency in the 0 to 6 s window following event offset. Each point reflects a participant's centre of mass for responses in the respective window, grouped by noise condition. Error bars represent ± 1 SEM across participants. **D.** Predicted probability of boundary detection as a function of NBSI (i.e., proportion of GPT-4 iterations identifying a boundary) for each noise condition. Solid lines show the logistic mixed effects model predictions; shaded bands represent the 95% confidence intervals ($n = 36$).

that listeners may rather be responding to the conclusion of the current event. In what follows, we thus focus our analyses on the proportion of responses time locked to event offset.

3.2.3.2. Effects of background noise on time locked responses. To evaluate how background noise affected the timing and magnitude of event boundary detection, we conducted an analysis that included baseline corrected event boundary responses in the post-boundary window. When time locked to event offset, results showed that the proportion of identified boundaries was significantly greater in the +2 dB SNR than the -4 dB SNR condition ($t_{67} = 2.72$, $p_T = 2.26 \times 10^{-2}$; main effect of noise condition: $F_{2,67.52} = 4.14$, $p = 2.02 \times 10^{-2}$, $\eta_p^2 = 0.11$, 95% CI [0.00, 0.25]; Fig. 4B), but comparable to the clear condition ($t_{67.1} = 0.58$, $p_T = 0.833$). Baseline corrected proportions in the clear condition did not differ from the -4 dB SNR condition after post hoc correction ($t_{67.1} = 2.12$, $p_T = 9.44 \times 10^{-2}$). Critically, identification of event boundaries was significant event at -4 dB SNR, where participants missed or misheard more than 30% of words (Fig. 4A), as indicated by the significant difference between the pre- and post-boundary windows ($t_{169} = 2.65$, $p = 8.73 \times 10^{-3}$, $d = 0.62$, 95% CI [0.07, 1.18]).

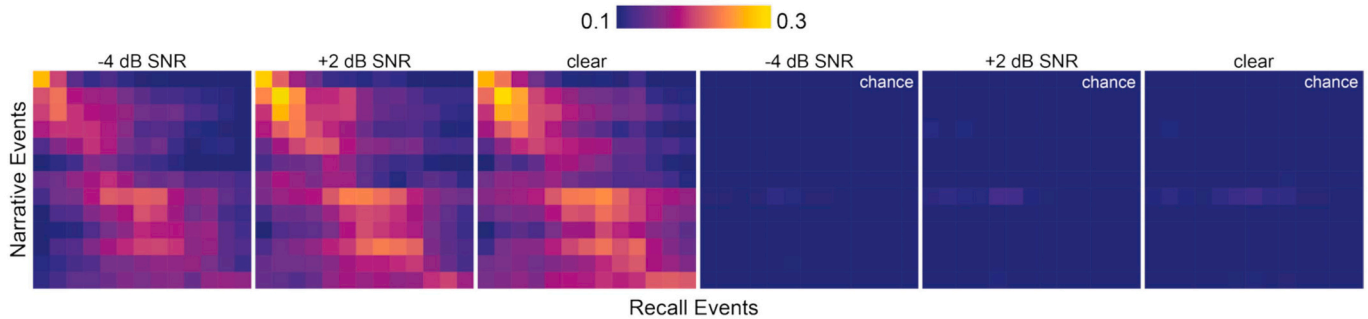
Because increased background noise may disrupt the ability for participants to identify normative event boundaries, we subsequently assessed whether segmentation responses occur later as background noise increases (visually suggested by the time courses in Fig. 4A). Latency was computed as a weighted average time of boundary responses within a 0 to 6 s window after event offset. Results did not indicate a

significant effect of background noise on responses latency ($F_{2,59.24} = 2.41$, $p = 9.83 \times 10^{-2}$, $\eta_p^2 = 0.08$, 95% CI [0.00, 0.21], Fig. 4C). While the pattern suggests a trend toward delayed responses after event offset in the -4 dB SNR condition, the evidence did not reach statistical significance.

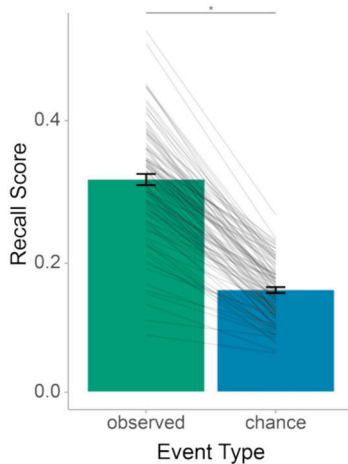
3.2.4. Normative boundary salience index

To determine whether the salience of normative event boundaries is reflected in participants' responses, we computed a logistic mixed effect model with the probability of boundary identification as the outcome variable. The continuous predictor included the normative boundary salience index, defined as the proportion of LLM iterations that identified a given normative boundary. The results revealed significant main effects of salience ($\chi^2_1 = 13.53$, $p = 2.34 \times 10^{-4}$) and noise condition ($\chi^2_2 = 8.07$, $p = 1.77 \times 10^{-2}$); however, the interaction between the salience and noise conditions was not significant ($\chi^2_2 = 3.81$, $p = .149$). A follow-up analysis of the estimated trends indicated the normative boundary salience index positively predicted boundary identification in the +2 dB SNR ($\beta = 2.13$, $z = 3.64$, $p = 2.72 \times 10^{-4}$), with weaker and non-significant slopes in the -4 dB SNR ($\beta = 0.57$, $z = 1.00$, $p = .319$) and clear ($\beta = 1.03$, $z = 1.84$, $p = 6.53 \times 10^{-2}$) conditions. Pairwise comparisons of these slopes did not reveal significant differences ($z = 1.90$, $p_T = 0.139$). The salience of event boundaries seems to drive event segmentation mainly under moderate noise (Fig. 4D).

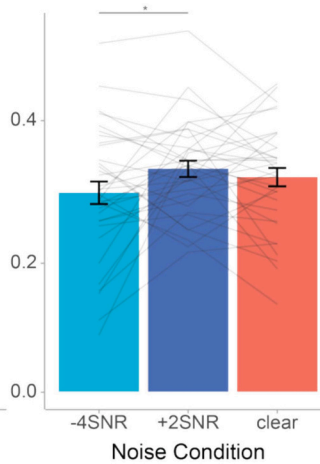
A Narrative × recall matrices



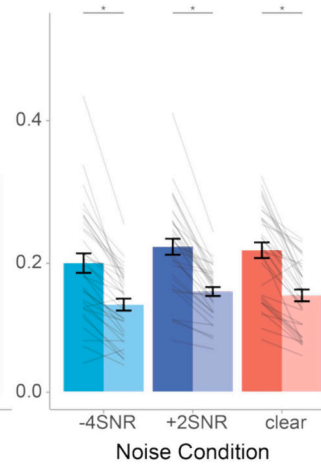
B Event recall vs chance



C Noise effect on recall



D Forward vs reverse recall



E Temporal divergence

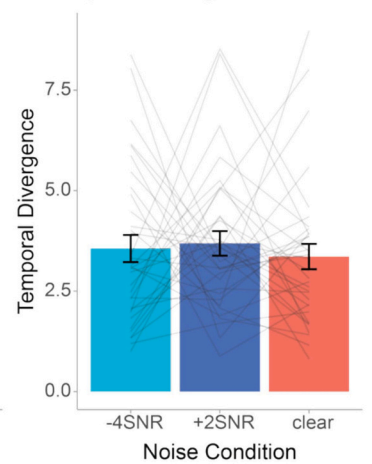


Fig. 5. Recall Results. A. Narrative × recall matrices and chance level matrices (derived from non-corresponding narratives). For visualization of the average matrices across narratives, matrices were transformed to square matrices based on the median number of narrative events. B. Event recall scores calculated for both corresponding and non-corresponding narratives. C. Recall scores across noise conditions. Lines represent individual participant's average event recall scores within narrative. D. Temporal divergence. Average recall scores along the forward diagonal (dark; indicating correct temporal sequence) are compared to scores along the reverse diagonal (light; indicating reversed sequence). Higher scores along the forward diagonal reflect more accurate temporal recall. E. Temporal divergence. Higher values reflect greater disruption in recall order. Error bars represent ± 1 SEM across participants ($n = 36$).

3.3. Event recall

Fig. 5A presents the narrative $\times$ recall similarity matrices for each noise condition alongside chance level matrices derived from non-corresponding narratives. These similarity matrices form the basis for the recall metrics reported below, with higher similarity values reflecting better recall for a given event, and a pronounced diagonal structure reflecting accurate temporal organization of the narrative.

3.3.1. Event recall scores

We first examined participant event recall performance by comparing recall scores for the corresponding narrative to those for two non-corresponding narratives, invariant of noise conditions. As expected, event recall scores were higher when participants were evaluated against the corresponding narrative compared to the two other non-corresponding narratives ($F_{1,33} = 809.79$, $p = 8.48 \times 10^{-25}$, $\eta_p^2 = 0.96$, 95% CI [0.93, 0.97]; Fig. 5B). A subsequent model was fit to determine whether background noise modulated the recall of the corresponding narrative. Recall performance in the +2 dB SNR condition was evidently greater than the -4 dB SNR condition ($t_{32.1} = 2.46$, $p_T = 4.98 \times 10^{-2}$; main effect of noise condition: $F_{2,29.17} = 3.43$, $p = 4.60 \times 10^{-2}$, $\eta_p^2 = 0.19$, 95% CI [0.00, 0.41]; Fig. 5C). There were no significant differences in recall performance among the +2 dB SNR and clear conditions ($t_{31.8} = 1.18$, $p_T = 0.474$), and the -4 dB SNR and clear conditions ($t_{32.1} = 1.08$, $p_T = 0.531$). These results indicate that severe background noise can ultimately hinder recall performance.

3.3.2. Temporal order

Linear mixed effects analyses revealed that event recall scores along the forward diagonal of the recall matrix were greater than the reverse order scores across all noise conditions (smallest $t_{2883} = 9.00$, $p_T = 3.83 \times 10^{-19}$, main effect of temporality: $F_{1,2882.8} = 277.75$, $p = 1.31 \times 10^{-59}$, $\eta_p^2 = 0.09$, 95% CI [0.07, 0.11]; Fig. 5D). The interaction between noise condition and temporality was insignificant ($F_{2,2882.8} = 0.17$, $p = .840$, $\eta_p^2 = 3.40 \times 10^{-3}$, 95% CI [0.00, 0.01]), suggesting that narratives are meaningfully recalled in temporal order across noise conditions. A follow-up analysis using the temporal divergence metric similarly showed no significant effect of noise condition ($F_{2,64.02} = 0.43$, $p = .649$, $\eta_p^2 = 0.01$, 95% CI [0.00, 0.09]; Fig. 5E). Participants tended to recall events in the same temporal order as they occurred in the narrative, suggesting that the temporal organization of memory remains stable even under reduced speech clarity.

3.3.3. Segmentation performance on recall

To examine the unique contribution of segmentation beyond noise on recall performance, we residualized (García et al., 2020; Geldhof et al., 2013; Perin et al., 2014) segmentation scores on noise condition and subsequently modeled recall as a function of noise condition, segmentation residuals, and their interaction. When using the agreement index, results revealed a significant main effect of noise condition ($F_{2,1433.2} = 5.22$, $p = 5.48 \times 10^{-3}$, $\eta_p^2 = 7.24 \times 10^{-3}$, 95% CI [0.00, 0.02]; Fig. 6A), segmentation ($F_{1,1441.9} = 10.65$, $p = 1.13 \times 10^{-3}$, $\eta_p^2 = 7.33 \times 10^{-3}$, 95% CI [0.00, 0.02]), and their interaction ($F_{2,1330.0} = 4.12$, $p = 1.65 \times 10^{-2}$, $\eta_p^2 = 6.16 \times 10^{-3}$, 95% CI [0.00, 0.02]). A follow-up analysis of the estimated trends indicated that the agreement index positively predicted recall performance in the +2 dB SNR ($\beta = 0.36$, $t_{3.51} = 3.51$, $p = 4.58 \times 10^{-4}$), with weaker and non-significant slopes in the -4 dB SNR ($\beta = -9.22 \times 10^{-3}$, $t_{1378.71} = 0.11$, $p = .909$) and clear conditions ($\beta = 4.13 \times 10^{-2}$, $t_{1384.47} = 0.57$, $p = .570$). Pairwise comparisons of these slopes revealed significant differences between the +2 dB SNR and both the -4 dB SNR ($t_{1366.18} = 2.69$, $p_T = 1.96 \times 10^{-2}$) and clear conditions ($t_{1350.48} = 2.39$, $p_T = 4.50 \times 10^{-2}$). No differences between the -4 dB SNR and clear conditions were observed ($t_{1276.07} = 0.41$, $p_T = 0.912$).

When the same analysis was performed used the post-boundary amplitudes from the event-offset analysis, the main effect of noise condition remained ($F_{2,1411.97} = 5.60$, $p = 3.78 \times 10^{-3}$, $\eta_p^2 = 7.87 \times 10^{-3}$, 95% CI [0.00, 0.02]; Fig. 6B); however, neither segmentation ($F_{1,1414.15} = 0.19$, $p = .667$, $\eta_p^2 = 1.31 \times 10^{-4}$, 95% CI [0.00, 0.00]) nor their interaction ($F_{2,1263.46} = 0.19$, $p = .829$, $\eta_p^2 = 2.97 \times 10^{-4}$, 95% CI [0.00, 0.00]) reached significance. In other words, while segmentation aligned with group-level boundary structure predicted recall beyond noise, the proportion of normative boundaries identified did not explain addition variance in recall performance.

3.3.4. Event centrality

In the final analysis, we examined whether the importance or centrality of an event may similarly explain memory performance. To explore this, we examined the relationship between recall and the centrality of each event, defined as the degree to which an event is semantically connected to other events within the narrative. A significant effect of centrality was evident ($F_{1,1286.52} = 133.58$, $p = 1.82 \times 10^{-19}$, $\eta_p^2 = 0.09$, 95% CI [0.07, 0.13]; Fig. 6C), indicating that more central events are more likely to be recalled with greater accuracy. Estimated slopes indicate that this effect was present in the -4 dB SNR

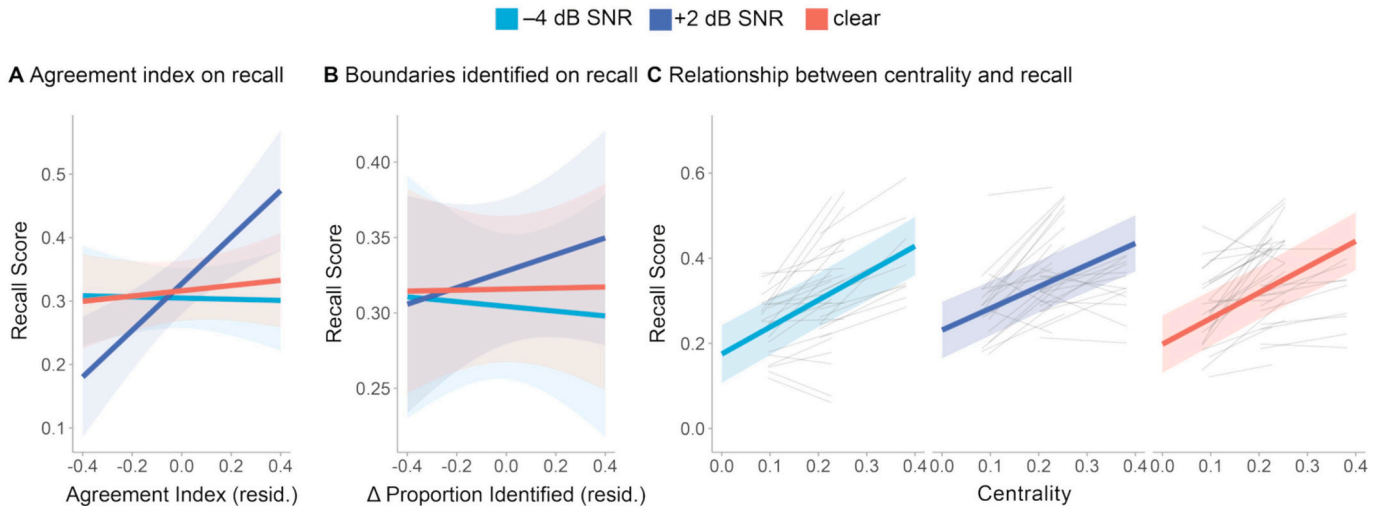


Fig. 6. Narrative Structure to Recall Relationship. A. Predicted event recall scores as a function of residualized agreement index. Solid lines show the model predictions; shaded bands represent the 95% confidence intervals. B. Predicted event recall scores as a function of the baseline corrected proportion of events identified in the post-boundary (1 to 3 s) windows. Solid lines show the model predictions; shaded bands represent the 95% confidence intervals. C. Event centrality predictions for recall scores. Coloured lines represent model predicted trends, with the shaded region indicating the 95% confidence intervals ($n = 36$).

($t_{1121.74} = 7.68$, $p_T = 3.45 \times 10^{-14}$), +2 dB SNR ($t_{1102.11} = 5.87$, $p_T = 5.85 \times 10^{-9}$), and clear conditions ($t_{1060.35} = 7.93$, $p_T = 5.61 \times 10^{-15}$). The interaction between centrality and noise condition was not significant ($F_{2,648.85} = 1.61$, $p = .200$, $\eta_p^2 = 4.94 \times 10^{-3}$, 95% CI [0.00, 0.02]). Specifically, the centrality effect presented similarity across all noise conditions (largest $\beta = 0.15$, $t_{614} = 1.59$, $p_T = 0.250$). Collectively, these findings indicate that event centrality can support the encoding of events even in difficult listening environments.

4. Discussion

The current study investigated the impact of background noise on the perceptual organization and encoding of spoken narrative events. Beyond demonstrating that adverse listening conditions impair comprehension and downstream memory, our data shows that participants continue to segment narratives into meaningful events despite a ~30% reduction in intelligibility in high levels of background noise (−4 dB SNR). At the same time, normative event boundaries emerged as particularly salient under moderate masking, suggesting that listeners may rely more heavily on narrative structures under moderately challenging conditions. The accuracy with which listeners recalled a narrative decreased for more difficult listening conditions, as expected; however, our data also suggests that listeners that show better segmentation agreement also recall narrative events more accurately, specifically under moderate background noise. These findings contribute to the growing body of literature emphasizing the relationship between event segmentation in supporting encoding and provides novel insights into how auditory experiences shape comprehension.

4.1. Cognitive mechanisms supporting segmentation of continuous speech

Consistent with previous literature and with predictions from the event segmentation theory (Zacks et al., 2007) and prior behavioural work (Baldassano et al., 2017; Kurby & Zacks, 2011; Speer & Zacks, 2005), we found that listeners reliably identified event boundaries in a manner that reflected the innate narrative structure even under naturalistic listening conditions. The ability to identify events in continuous speech highlights that segmentation is guided by conceptual understanding and is supported by situation models that integrate semantic, contextual, and potentially prosodic information. This replicates a robust literature demonstrating that event boundaries correspond to moments of significant change in situational context resulting in model updating (Magliano et al., 2001; Rinck & Bower, 2000; Zacks et al., 2009; Zwaan, 1996; Zwaan & Radvansky, 1998), suggesting that listeners continuously track ongoing structure even when perceptual details vary.

The current work went beyond traditional event segmentation metrics (Sasmita & Swallow, 2022) to better understand how individuals perceive normative event boundaries. Standard metrics, such as the agreement index (Kurdy & Zacks, 2011; Sasmita & Swallow, 2022; Zacks et al., 2006) assess segmentation performance as interparticipant agreement (Sasmita & Swallow, 2022; Zacks et al., 2006). Such approaches limit the ability to assess individual differences, as they rely on group-based comparisons rather than external benchmarks. To address these limitations, we introduced a novel time locked event boundary analysis, which evaluates participant responses with event boundaries identified independently by a large language model (LLM; Michelmann et al., 2025; Panella et al., 2025). This enabled us to not only determine whether a participant identified an event boundary, but precisely when they did so. The analyses revealed that participants' segmentation patterns were temporally aligned with the LLM-identified normative event boundaries rather than event centres, providing additional validation in the utility of LLMs in identifying event boundaries (Fig. 3).

Our time locked analysis may also enable us to evaluate the key assumption of the EST, which dictates the foundations in prediction error. Specifically, the model emphasizes that when the predictions no

longer align with the current experience, individuals perceive an event boundary and update their event model (Nolden et al., 2024; Reynolds et al., 2007; Zacks et al., 2007, 2011). Our time locked analysis enabled us to differentiate between reactive segmentation, in which boundaries are identified only after perceived input conflicts with expectations, and anticipatory segmentation, in which listeners preemptively commit to an event boundary before sufficient new contextual information is available to generate a prediction error. The proportion of identified boundaries peaked within one second of event offset, which includes decision and response time (Fig. 3). This likely does not provide enough time nor information for a context driven prediction error, suggesting that listeners may often anticipate an impending transition prior to new narrative content sufficiently unfolding. This temporal alignment may support an anticipatory strategy of segmentation. Recent computational accounts have proposed that event boundaries perception may also arise from fluctuations in prediction uncertainty – moments when confidence in ongoing predictions decreases – rather than prediction error alone (Franklin et al., 2020; Nguyen et al., 2024; Smith et al., 2025). From this perspective, the anticipatory responses observed here may reflect increases in uncertainty as listeners detect cues indicating a potential emergence of novel narrative information, prompting them to prepare for an update before a contextual violation occurs. This framework unites reactive and proactive views of event segmentation (Franklin et al., 2020; Nguyen et al., 2024; Reynolds et al., 2007; Zacks et al., 2007, 2011), emphasizing that narrative and speech comprehension may involve not only responding to prediction errors but also continuously monitoring uncertainty as information continuously unfolds. This anticipatory pattern may also reflect the broader complexity of speech comprehension. Speech materials provide not only semantic and contextual information, but also prosodic and acoustic cues (Bradlow & Alexander, 2007; Brown et al., 2015; McGettigan et al., 2012; Nygaard et al., 2009). As spatial changes (i.e., fade to black in film) often trigger an event boundary in the visual domain (Boltz, 1992; Kurby & Zacks, 2008; Newtonson & Engquist, 1976; Schwan et al., 2000; Zacks et al., 2009), these auditory features may also coincide with meaningful transitions in speech. It is worth investigating in future research, the specific contributions that acoustic cues may provide to event cognition.

4.2. Event segmentation under challenging listening conditions

Critically, the current study examined how event segmentation during listening is affected by the background noise – a proxy for listening effort. We used background noise levels that reduced speech intelligibility by approximately 10% (+2 dB SNR) and 30% (−4 dB SNR) relative to speech-in-quiet. This reduction in speech intelligibility is consistent with previous work (Irsik et al., 2022a, 2022b; Kalikow et al., 1977; Moore & Glasberg, 1993).

When narratives were presented under challenging acoustics (i.e., −4 dB SNR), participants identified fewer event boundaries and showed reduced agreement in boundary placement (Fig. 2). When assessing participant responses time locked to normative event boundaries, it was also evident that segmentation responses in the −4 dB SNR were attenuated and occurred slightly later in time relative to the clear and +2 dB SNR conditions (Fig. 4C). Although the latency differences did not reach significance, the numerical trend aligns with the assumptions that greater processing demands may delay downstream cognitive processes (Fitousi & Wenger, 2011; Lavie et al., 2004; Zekveld et al., 2011). Critically, even when participants missed or misheard approximately 30% of words, they were able to identify event boundaries above baseline levels (Fig. 4B), suggesting they were able to follow the narrative thread despite adverse listening conditions.

Segmentation around normative boundaries appeared somewhat more pronounced in the +2 dB SNR condition, suggesting that moderate background noise did not disrupt segmentation, but may rather emphasize highly salient narrative transition to guide event

organization, reflecting increased engagement or a more strategic allocation of attentional resources. This aligns with previous research showing that moderate background noise does not necessarily reduce narrative absorption (Herrmann & Johnsrude, 2020b) nor engagement while listening (Irsik et al., 2022b).

Together, these results suggest that perceptual segmentation becomes less reliable when speech is masked by background babble at high levels, reflecting increased listening effort under conditions of greater acoustic degradation. Although both levels of background noise involve energetic masking, the -4 dB SNR condition produces substantially greater reductions in speech intelligibility (Mattys et al., 2012). This increases the demand on attentional and cognitive resources required to maintain a coherent representation of the narrative. Nonetheless, participants retained moderate intelligibility ($\sim 30\%$ reduction from clear speech) and continued to identify meaningful event boundaries even under this condition, highlighting the robustness of higher-level narrative processing regardless of substantial bottom-up degradation. It is also worth considering that the experimental task required participants to continuously comprehend the narrative while concurrently indicating perceived event boundaries, which is an inherent dual-task feature of behavioural segmentation paradigms (Kurby & Zacks, 2011; Sargent et al., 2013; Sasmita & Swallow, 2022; Zacks et al., 2001, 2006) and may contribute to overall cognitive demands under more challenging listening conditions; however, this task structure was constant across all conditions, suggesting that observed differences are driven primarily by changes in listening demands rather than task requirements. Consistent with this, prior research has shown that listening effort impairs working memory (Mattys et al., 2012; Rönnberg et al., 2013) and the maintenance of high-level cognitive representations such as event models (Zwaan & Radvansky, 1998). In contrast, there were no evident differences in segmentation performance and timing between the clear and $+2$ dB SNR conditions, indicating that event organization remains stable even when intelligibility is moderately reduced.

4.3. Event recall and preservation of narrative structure following encoding in noise

Listeners demonstrated reliable memory for narrative content across all listening conditions. Recall performance exceeded chance levels and participants reliably recalled events in the order they were presented. This preservation of temporal order was stable across noise conditions, suggesting that the structural organization of narrative events at encoding are maintained even when the intelligibility of speech is compromised. This aligns with prior work showing that individuals have the tendency to recall events in a forward, chronological direction (Baldassano et al., 2017; J. Chen et al., 2017; Dougherty et al., 2023; Healey, 2018; Heusser et al., 2021; Wen & Egner, 2022). The fact that participants could recall narratives above chance despite losing approximately 30% of words in the most adverse listening conditions underscores the robustness of gist-based encoding over verbatim recall (Hasher & Clifton, 1974; Mehler, 1963; Mehler & Miller, 1964; Sachs, 1967). The ability to reconstruct meaningful narratives even from incomplete input highlights recall as an indicator of successful communication and comprehension in real-world listening.

Despite the robust temporal structure, we observed that the quality of event recall varied with background noise. Recall accuracy for narrative events was greater in the $+2$ dB SNR than in the -4 dB SNR condition, aligning with the intuition that increased background noise can impair encoding of spoken materials. The additional cognitive resource required to comprehend masked speech comes at a cost of processing resources that would otherwise be available for encoding (McCoy et al., 2005; Peelle, 2018). Numerically, recall performance was greatest for the moderate noise condition ($+2$ dB SNR), and recall of clear speech did not differ between either speech-in-noise condition (Fig. 5C). This aligns with the literature on *desirable difficulty*, where moderate levels of challenge can enhance processing and retention

(Bjork & Bjork, 2011; Vogel & Schwabe, 2016). Segmentation ability (i. e., agreement index) predicted recall performance specifically at the $+2$ dB SNR, not in the clear nor -4 dB SNR condition (Fig. 6A). These findings extend earlier evidence that segmentation ability predicts memory (Nolden et al., 2024; Sargent et al., 2013; Smith et al., 2024) by showing that the relationship is more apparent under conditions that engage sustained attention. The current results indicate that rather than degrading comprehension linearly, moderate challenges may promote attention and engagement (Herrmann & Johnsrude, 2020b), potentially contributing to deeper encoding of the narrative. When considered alongside the preserved temporal order, these findings indicate that even when fewer event details are encoded under noise, listeners retain a coherent narrative structure. Such resilience likely reflects the influence of narrative schemas – internal models of how narratives typically unfold – which guide memory reconstruction and help preserve the sequencing of events encoded during listening (Abelson, 1981; van Dijk & Kintsch, 1983; Lee & Chen, 2022; Zwaan & Radvansky, 1998).

Our findings also reveal a consistent relationship between event structure and memory through recently developed centrality metrics (Lee & Chen, 2022). Events that were more central to the narrative, were more likely to be recalled, regardless of listening condition (Fig. 6C). This aligns with previous work suggesting that centrality serves as an anchor which helps to enhance the accessibility of important moments during retrieval (Lee & Chen, 2022; Trabasso & Sperry, 1985; Trabasso & Van Den Broek, 1985). Listeners may engage in selective encoding (Uncapher & Rugg, 2009; Wright et al., 1977), where semantically important content is more deeply processed (Craik & Lockhart, 1972; Craik & Tulving, 1975). Recent work suggests that centrality may capture more than surface-level features but rather reflects deeper situational models (Zwaan et al., 1995; Zwaan & Radvansky, 1998) that not only support the encoding of individual events, but also represent the causal and semantic relationships that link those events across time (Lee & Chen, 2022). These central structures may ultimately serve as scaffolds for organizing memory, enabling listeners to retain meaningful temporal information even under adverse listening conditions.

5. Conclusion

The current experiment evaluated the relationship between the perceptual organization of speech and its downstream consequences for memory encoding. Using naturalistic spoken narratives, we found that listeners segment and recall events in a structured manner. Our results indicate that segmentation was primarily anticipatory, with listeners identifying event boundaries relative to the end of an event rather than at the beginning of a new event, which potentially challenges the prediction error framework of the event segmentation theory. Critically, event segmentation and recall were moderated by listening difficulty. Segmentation was less consistent, and memory performance declined for highly masked speech, whereas temporal order memory remained intact, suggesting that that overall narrative structure was preserved. We observed some indications that moderate background noise enhances event processing and recall, potentially through attentional engagement. The current results provide novel insights into the relationship between background noise and event processing, highlighting potential mechanisms through which listening effort may contribute to perceptual and memory failures in everyday communication.

CRedit authorship contribution statement

Ryan A. Panella: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alexander J. Barnett:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. **Yulia Lamekina:** Writing – review & editing. **Morgan D. Barense:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **Björn Herrmann:**

Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ryan Panella reports financial support was provided by Canadian Institutes of Health Research. Björn Herrmann reports financial support was provided by Canadian Institutes of Health Research. Björn Herrmann reports financial support was provided by Natural Sciences and Engineering Research Council of Canada. Björn Herrmann reports financial support was provided by Canada Research Chairs Program. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Tiffany Lao for her help with data collection; Sarah Bobbitt, Saba Junaid, and Andrew Cole for their help with manual transcribing of recall data; and Joshua Koh for his thoughtful insight. This research was supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant: RGPIN-2021-02602), Canadian Institutes of Health Research (Funding Reference Number – RAP: 193310, BH: 195994), and the Canada Research Chairs Program (CRC-2023-00383).

Data availability

Data and analytical code are available via GitHub at github.com/ryanapanella/EventNoise.

References

- Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist*, 36(7), 715–729. <https://doi.org/10.1037/0003-066X.36.7.715>
- Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. J. (2017). Self-Reported Listening-Related Effort and Fatigue in Hearing-Impaired Adults. *Ear and Hearing*, 38(1), e39–e48. <https://doi.org/10.1097/AUD.0000000000000361>
- Allison, K. M. (2020). Measuring Speech Intelligibility in Children With Motor Speech Disorders. *Perspectives of the ASHA Special Interest Groups*, 5(4), 809–820. https://doi.org/10.1044/2020_PERSP-19-00110
- Allison, K. M., Millington, M., & Grimm, A. (2025). Can We Trust Our Ears? How Accurate and Reliable Are Speech-Language Pathologists' Estimates of Children's Speech Intelligibility? *American Journal of Speech-Language Pathology*, 34(2), 853–867. https://doi.org/10.1044/2024_AJSLP-24-00247
- Baddeley, A. D., & Hitch, G. (1974). Working Memory. In , 8. *Psychology of Learning and Motivation* (pp. 47–89). Elsevier. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- 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>
- Barker, R. G. (Ed.). (1963). *The stream of behavior: Explorations of its structure & content*. Appleton-Century-Crofts. <https://doi.org/10.1037/11177-000>
- Barker, R. G., & Wright, H. F. (1954). *Midwest and its children: The psychological ecology of an American town*. Row, Peterson and Company. <https://doi.org/10.1037/10027-000>
- Barnett, A. J., Nguyen, M., Spargo, J., Yadav, R., Cohn-Sheehy, B. I., & Ranganath, C. (2024). Hippocampal-cortical interactions during event boundaries support retention of complex narrative events. *Neuron*, 112(2), 319–330.e7. <https://doi.org/10.1016/j.neuron.2023.10.010>
- Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00328>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Ben-Shachar, M., Lüdtke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *The Journal of Open Source Software*, 5(56), Article 2815. <https://doi.org/10.21105/joss.02815>
- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a Test of Speech Perception in Noise. *Journal of Speech, Language, and Hearing Research*, 27(1), 32–48. <https://doi.org/10.1044/jshr.2701.32>
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.
- Boltz, M. (1992). Temporal accent structure and the remembering of filmed narratives. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1), 90–105. <https://doi.org/10.1037/0096-1523.18.1.90>
- Borrie, S. A., Barrett, T. S., & Yoho, S. E. (2019). Autoscoring: An open-source automated tool for scoring listener perception of speech. *The Journal of the Acoustical Society of America*, 145(1), 392–399. <https://doi.org/10.1121/1.5087276>
- Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, 121(4), 2339–2349. <https://doi.org/10.1121/1.2642103>
- Brown, M., Salverda, A. P., Gunlogson, C., & Tanenhaus, M. K. (2015). Interpreting prosodic cues in discourse context. *Language, Cognition and Neuroscience*, 30(1–2), 149–166. <https://doi.org/10.1080/01690965.2013.862285>
- Burunat, I., Levitin, D. J., & Toivainen, P. (2024). Breaking (musical) boundaries by investigating brain dynamics of event segmentation during real-life music-listening. *Proceedings of the National Academy of Sciences of the United States of America*, 121(36), Article e2319459121. <https://doi.org/10.1073/pnas.2319459121>
- Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Céspedes, M., Yuan, S., Tar, C., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder for English. In E. Blanco, & W. Lu (Eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 169–174). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D18-2029>
- Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2025). Unleashing the potential of prompt engineering for large language models. *Patterns*, 6(6), Article 101260. <https://doi.org/10.1016/j.patter.2025.101260>
- 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>
- Clayson, P. E., Baldwin, S. A., & Larson, M. J. (2013). How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study. *Psychophysiology*, 50(2), 174–186. <https://doi.org/10.1111/psyp.12001>
- Corrente, M., & Bourgeault, I. (2022). *Innovation in Transcribing Data: Meet Otter.ai*. SAGE Publications, Ltd. <https://doi.org/10.4135/9781529799033>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [https://doi.org/10.1016/s0022-5371\(72\)80001-x](https://doi.org/10.1016/s0022-5371(72)80001-x)
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268–294. <https://doi.org/10.1037/0096-3445.104.3.268>
- Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31(2), 218–236. [https://doi.org/10.1016/0749-596X\(92\)90012-M](https://doi.org/10.1016/0749-596X(92)90012-M)
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 113–121. <https://doi.org/10.1037/0096-1523.14.1.113>
- Davis, E. E., & Campbell, K. L. (2022). Event boundaries structure the contents of long-term memory in younger and older adults. *Memory*, 31(1), 47–60. <https://doi.org/10.1080/09658211.2022.2122998>
- Davis, M. H., & Johnsruide, I. S. (2003). Hierarchical Processing in Spoken Language Comprehension. *The Journal of Neuroscience*, 23(8), 3423–3431. <https://doi.org/10.1523/JNEUROSCI.23-08-03423.2003>
- Dickman, H. R. (1963). The Perception of Behavioral Units. In R. G. Barker (Ed.), *The stream of behavior: Explorations of its structure & content* (pp. 23–41). Appleton-Century-Crofts. <https://doi.org/10.1037/11177-002>
- Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the late positive complex: Mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology*, 41(5), 665–678. <https://doi.org/10.1111/j.1469-8986.2004.00193.x>
- Dougherty, M. R., Halpern, D., & Kahana, M. J. (2023). Forward and backward recall dynamics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 49(11), 1752–1772. <https://doi.org/10.1037/xlm0001254>
- Ezzyat, Y., & Davachi, L. (2011). What Constitutes an Episode in Episodic Memory? *Psychological Science*, 22(2), 243–252. <https://doi.org/10.1177/0956797610393742>
- Fitoussi, D., & Wenger, M. J. (2011). Processing capacity under perceptual and cognitive load: A closer look at load theory. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 781–798. <https://doi.org/10.1037/a0020675>
- Flanagan, J. L. (1972). *Speech Analysis Synthesis and Perception*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-01562-9>
- Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (Third). SAGE Publications, Ltd. <https://www.john-fox.ca/Companion/>
- 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. *Psychological Review*, 127(3), 327–361. <https://doi.org/10.1037/rev0000177>
- García, C. B., Salmerón, R., García, C., & García, J. (2020). Residualization: Justification, properties and application. *Journal of Applied Statistics*, 47(11), 1990–2010. <https://doi.org/10.1080/02664763.2019.1701638>
- Geldhof, G. J., Pornprasertmanit, S., Schoemann, A. M., & Little, T. D. (2013). Orthogonalizing Through Residual Centering: Extended Applications and Caveats.

- Educational and Psychological Measurement, 73(1), 27–46. <https://doi.org/10.1177/0013164412445473>
- Goh, C. M., & Wallace, M. (2018). Lexical Segmentation in Listening. In J. I. Lontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (1st ed., pp. 1–7). Wiley. <https://doi.org/10.1002/9781118784235.eelt0603>
- Hasher, L., & Clifton, D. (1974). A developmental study of attribute encoding in free recall. *Journal of Experimental Child Psychology*, 17(2), 332–346. [https://doi.org/10.1016/0022-0965\(74\)90077-0](https://doi.org/10.1016/0022-0965(74)90077-0)
- Healey, M. K. (2018). Temporal contiguity in incidentally encoded memories. *Journal of Memory and Language*, 102, 28–40. <https://doi.org/10.1016/j.jml.2018.04.003>
- Herrmann, B. (2025a). Language-agnostic, automated assessment of listeners' speech recall using large language models. *Trends in Hearing*, 29, Article 23312165251347131. <https://doi.org/10.1177/23312165251347131>
- Herrmann, B. (2025b). Leveraging natural language processing models to automate speech-intelligibility scoring. *Speech, Language and Hearing*, 28(1), Article 2374160. <https://doi.org/10.1080/2050571X.2024.2374160>
- Herrmann, B., & Johnsrude, I. S. (2020b). Absorption and enjoyment during listening to acoustically masked stories. *Trends in Hearing*, 24, Article 233121652096785. <https://doi.org/10.1177/233121652096785>
- Herrmann, B., & Johnsrude, I. S. (2020a). A model of listening engagement (MoLE). *Hearing Research*, 397, Article 108016. <https://doi.org/10.1016/j.heares.2020.108016>
- Heusser, A. C., Fitzpatrick, P. C., & Manning, J. R. (2021). Geometric models reveal behavioural and neural signatures of transforming experiences into memories. *Nature Human Behaviour*, 5(7), 905–919. <https://doi.org/10.1038/s41562-021-01051-6>
- Hodgson, M., Steininger, G., & Razavi, Z. (2007). Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments. *The Journal of the Acoustical Society of America*, 121(4), 2023–2033. <https://doi.org/10.1121/1.2535571>
- Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, 46(3), 269–299. <https://doi.org/10.1006/jmps.2001.1388>
- Hustad, K. C., Oakes, A., & Allison, K. (2015). Variability and diagnostic accuracy of speech intelligibility scores in children. *Journal of Speech, Language, and Hearing Research : JSLHR*, 58(6), 1695–1707. <https://doi.org/10.1044/2015.JSLHR-S-14-0365>
- Irisk, V. C., Johnsrude, I. S., & Herrmann, B. (2022a). Age-related deficits in dip-listening evident for isolated sentences but not for spoken stories. *Scientific Reports*, 12(1), Article 5898. <https://doi.org/10.1038/s41598-022-09805-6>
- Irisk, V. C., Johnsrude, I. S., & Herrmann, B. (2022b). Neural Activity during Story Listening Is Synchronized across Individuals Despite Acoustic Masking. *Journal of Cognitive Neuroscience*, 34(6), 933–950. <https://doi.org/10.1162/jocn.a.01842>
- Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, 39(3), 569–591. <https://doi.org/10.1017/S0142716417000534>
- Kallick, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, 61(5), 1337–1351. <https://doi.org/10.1121/1.381436>
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. <https://doi.org/10.1126/science.1152408>
- Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward Phonetic Intelligibility Testing in Dysarthria. *Journal of Speech and Hearing Disorders*, 54(4), 482–499. <https://doi.org/10.1044/jshd.5404.482>
- Kenward, M. G., & Roger, J. H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics*, 53(3), 983–997. <https://doi.org/10.2307/2533558>
- Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, 45, 326–347. <https://doi.org/10.1016/j.csl.2017.01.005>
- Kumar, M., Goldstein, A., Michelmann, S., Zacks, J. M., Hasson, U., & Norman, K. A. (2023). Bayesian surprise predicts human event segmentation in story listening. *Cognitive Science*, 47(10), Article e13343. <https://doi.org/10.1111/cogs.13343>
- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2), 72–79. <https://doi.org/10.1016/j.tics.2007.11.004>
- Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. *Memory & Cognition*, 39(1), 75–91. <https://doi.org/10.3758/s13421-010-0027-2>
- Kurby, C. A., & Zacks, J. M. (2018). Preserved neural event segmentation in healthy older adults. *Psychology and Aging*, 33(2), 232–245. <https://doi.org/10.1037/pag0000226>
- 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>
- Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General*, 133(3), 339–354. <https://doi.org/10.1037/0096-3445.133.3.339>
- 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>
- Lenth, R. V. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.10.1) [Computer software]. <https://CRAN.R-project.org/package=emmeans>
- Liu, S., Nakajima, Y., Chen, L., Arndt, S., Kakizoe, M., Elliott, M. A., & Remijn, G. B. (2022). How Pause Duration Influences Impressions of English Speech: Comparison Between Native and Non-native Speakers. *Frontiers in Psychology*, 13, Article 778018. <https://doi.org/10.3389/fpsyg.2022.778018>
- Lunner, T., & Sundewall-Thoren, E. (2007). Interactions between Cognition, Compression, and Listening Conditions: Effects on Speech-in-Noise Performance in a Two-Channel Hearing Aid. *Journal of the American Academy of Audiology*, 18(7), 604–617. <https://doi.org/10.3766/jaaa.18.7.7>
- Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, 15(5), 533–545. <https://doi.org/10.1002/acp.724>
- Mansour, N., Marschall, M., May, T., Westermann, A., & Dau, T. (2021). A method for realistic, conversational signal-to-noise ratio estimation. *The Journal of the Acoustical Society of America*, 149(3), 1559–1566. <https://doi.org/10.1121/10.0003626>
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language & Cognitive Processes*, 27(7–8), 953–978. <https://doi.org/10.1080/01690965.2012.705006>
- McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing Loss and Perceptual Effort: Downstream Effects on Older Adults' Memory for Speech. *The Quarterly Journal of Experimental Psychology Section A*, 58(1), 22–33. <https://doi.org/10.1080/02724980443000151>
- McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H., & Scott, S. K. (2012). Speech comprehension aided by multiple modalities: Behavioural and neural interactions. *Neuropsychologia*, 50(5), 762–776. <https://doi.org/10.1016/j.neuropsychologia.2012.01.010>
- Mehler, J. (1963). Some effects of grammatical transformations on the recall of english sentences. *Journal of Verbal Learning and Verbal Behavior*, 2(4), 346–351. [https://doi.org/10.1016/S0022-5371\(63\)80103-6](https://doi.org/10.1016/S0022-5371(63)80103-6)
- Mehler, J., & Miller, G. A. (1964). RETROACTIVE INTERFERENCE IN THE RECALL OF SIMPLE SENTENCES. *British Journal of Psychology*, 55(3), 295–301. <https://doi.org/10.1111/j.2044-8295.1964.tb00913.x>
- Mei, Y., Chen, F., & Chen, X. (2024). Effects of language proficiency on selective attention patterns at segmenting boundaries in English audio sentences. *Brain Sciences*, 14(12), Article 1204. <https://doi.org/10.3390/brainsci14121204>
- Michelmann, S., Kumar, M., Norman, K. A., & Toneva, M. (2025). Large language models can segment narrative events similarly to humans. *Behavior Research Methods*, 57(1), 39. <https://doi.org/10.3758/s13428-024-02569-z>
- Michelmann, S., Price, A. R., Aubrey, B., Strauss, C. K., Doyle, W. K., Friedman, D., ... Norman, K. A. (2021). Moment-by-moment tracking of naturalistic learning and its underlying hippocampo-cortical interactions. *Nature Communications*, 12(1), 5394. <https://doi.org/10.1038/s41467-021-25376-y>
- Moore, B. C. J., & Glasberg, B. R. (1993). Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech. *The Journal of the Acoustical Society of America*, 94(4), 2050–2062. <https://doi.org/10.1121/1.407478>
- Newton, 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>
- Newton, 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](https://doi.org/10.1016/0022-1031(76)90076-7)
- Newton, 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>
- Nguyen, T. T., Bezdek, M. A., Gershman, S. J., Bobick, A. F., Braver, T. S., & Zacks, J. M. (2024). Modeling human activity comprehension at human scale: Prediction, segmentation, and categorization. *PNAS Nexus*, 3(10), Article pgae459. <https://doi.org/10.1093/pnasnexus/pgae459>
- Noah, T. (2016). *Born a crime: Stories from a South African childhood* (1st ed.). Spiegel & Grau.
- Nolden, S., Turan, G., Güler, B., & Günseli, E. (2024). Prediction error and event segmentation in episodic memory. *Neuroscience & Biobehavioral Reviews*, 157, Article 105533. <https://doi.org/10.1016/j.neubiorev.2024.105533>
- Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The Semantics of Prosody: Acoustic and Perceptual Evidence of Prosodic Correlates to Word Meaning. *Cognitive Science*, 33(1), 127–146. <https://doi.org/10.1111/j.1551-6709.2008.01007.x>
- Owoicho, P., Camp, J., & Kenter, T. (2024). *A Study of the Sensitivity of Subjective Listening Tests to Inter-sentence Pause Durations in English Speech*. Speech Prosody 2024, (pp. 462–466). <https://doi.org/10.21437/SpeechProsody.2024-94>
- Pandey, P. R., & Herrmann, B. (2025). The Influence of Semantic Context on the Intelligibility Benefit From Speech Glimpses in Younger and Older Adults. *Journal of Speech, Language, and Hearing Research*, 68(05), 2499–2516. <https://doi.org/10.1044/2025.JSLHR-24-00588>
- Panella, R. A. (2025). *EventRecall: Automated Event Recall Assessment (Version v1.0.0)* [Computer software]. Zenodo. <https://doi.org/10.5281/ZENODO.17467024>
- Panella, R. A., Barnett, A. J., Barens, M. D., & Herrmann, B. (2025). Event segmentation applications in large language model enabled automated recall assessments. *Communications Psychology*, 3(1), Article 184. <https://doi.org/10.1038/s44271-025-00359-7>
- Peelle, J. E. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear and Hearing*, 39(2), 204–214. <https://doi.org/10.1097/AUD.0000000000000494>
- Pearce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... 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>
- Perin, J., Preisser, J. S., Phillips, C., & Qaqish, B. (2014). Regression analysis of correlated ordinal data using orthogonalized residuals. *Biometrics*, 70(4), 902–909. <https://doi.org/10.1111/biom.12210>

- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., ... Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, 37(1), 5S–27S. <https://doi.org/10.1097/AUD.0000000000000312>
- R Core Team. (2024). R: A language and environment for statistical computing. (Version R 4.4.0) [Computer software]. R Foundation for Statistical Computing www.R-project.org/
- Racah, O., Chen, P., Gureckis, T. M., Poeppel, D., & Vo, V. (2024). *The Naturalistic Free Recall Dataset: Four stories, hundreds of participants, and high-fidelity transcriptions*. Center for Open Science. <https://doi.org/10.31234/osf.io/rxwqh>
- Racah, O., Doelling, K. B., Davachi, L., & Poeppel, D. (2023). Acoustic features drive event segmentation in speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 49(9), 1494–1504. <https://doi.org/10.1037/xlm0001150>
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108. <https://doi.org/10.1037/0033-295X.85.2.59>
- Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A Computational Model of Event Segmentation From Perceptual Prediction. *Cognitive Science*, 31(4), 613–643. <https://doi.org/10.1080/15326900701399913>
- Rinck, M., & Bower, G. H. (2000). Temporal and spatial distance in situation models. *Memory & Cognition*, 28(8), 1310–1320. <https://doi.org/10.3758/BF03211832>
- Rochester, S. R. (1973). The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research*, 2(1), 51–81. <https://doi.org/10.1007/BF01067111>
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20–27. <https://doi.org/10.1016/j.tics.2010.09.003>
- Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., ... Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, 7. <https://doi.org/10.3389/fnsys.2013.00031>
- Rusnock, C. F., & Bush, P. M. (2012). Case Study: An Evaluation of Restaurant Noise Levels and Contributing Factors. *Journal of Occupational and Environmental Hygiene*, 9(6), D108–D113. <https://doi.org/10.1080/15459624.2012.683716>
- Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2(9), 437–442. <https://doi.org/10.3758/BF03208784>
- Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., ... Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, 129(2), 241–255. <https://doi.org/10.1016/j.cognition.2013.07.002>
- Sasmita, K., & Swallow, K. M. (2022). Measuring event segmentation: An investigation into the stability of event boundary agreement across groups. *Behavior Research Methods*, 55(1), 428–447. <https://doi.org/10.3758/s13428-022-01832-5>
- Schwan, S., Garsoffky, B., & Hesse, F. W. (2000). Do film cuts facilitate the perceptual and cognitive organization of activity sequences? *Memory & Cognition*, 28(2), 214–223. <https://doi.org/10.3758/BF03213801>
- Shen, X., Houser, T., Smith, D. V., & Murty, V. P. (2023). Machine-learning as a validated tool to characterize individual differences in free recall of naturalistic events. *Psychonomic Bulletin & Review*, 30(1), 308–316. <https://doi.org/10.3758/s13423-022-02171-4>
- Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology*, 26(02), 183–196. <https://doi.org/10.3766/jaaa.26.2.7>
- Smith, M. E., Hall, C. S., Membreno, R., Quintero, D., & Zacks, J. M. (2024). Attention to event segmentation improves memory in young adults: A lifespan study. *Psychology and Aging*, 39(7), 750–769. <https://doi.org/10.1037/pag0000842>
- Smith, M. E., Zacks, J. M., & Reagh, Z. M. (2025). Events in the stream of behavior. *Current Opinion in Behavioral Sciences*, 65, Article 101581. <https://doi.org/10.1016/j.cobeha.2025.101581>
- 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>
- Swallow, K. M., Kemp, J. T., & Candan Simsek, A. (2018). The role of perspective in event segmentation. *Cognition*, 177, 249–262. <https://doi.org/10.1016/j.cognition.2018.04.019>
- Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event Boundaries in Perception Affect Memory Encoding and Updating. *Journal of Experimental Psychology: General*, 138(2), 236. <https://doi.org/10.1037/a0015631>
- Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, 24(5), 595–611. [https://doi.org/10.1016/0749-596X\(85\)90048-8](https://doi.org/10.1016/0749-596X(85)90048-8)
- Trabasso, T., & Van Den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24(5), 612–630. [https://doi.org/10.1016/0749-596X\(85\)90049-X](https://doi.org/10.1016/0749-596X(85)90049-X)
- Uncapher, M. R., & Rugg, M. D. (2009). Selecting for Memory? The Influence of Selective Attention on the Mnemonic Binding of Contextual Information. *Journal of Neuroscience*, 29(25), 8270–8279. <https://doi.org/10.1523/jneurosci.1043-09.2009>
- Van Der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., ... Yu, T. (2014). scikit-image: Image processing in Python. *PeerJ*, 2, Article e453. <https://doi.org/10.7717/peerj.453>
- Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.
- Rossum, G. v., & Drake, F. L. (2010). *The Python language reference* (Release 3.0.1 [Repr.]). Python Software Foundation.
- Vogel, S., & Schwabe, L. (2016). Learning and memory under stress: Implications for the classroom. *npj Science of Learning*, 1(1). <https://doi.org/10.1038/npjscilearn.2016.11>
- Wen, T., & Egner, T. (2022). Retrieval context determines whether event boundaries impair or enhance temporal order memory. *Cognition*, 225, Article 105145. <https://doi.org/10.1016/j.cognition.2022.105145>
- Wendt, D., Dau, T., & Hjortkjær, J. (2016). Impact of Background Noise and Sentence Complexity on Processing Demands during Sentence Comprehension. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00345>
- Wilson, R. H., McArdle, R., Watts, K. L., & Smith, S. L. (2012). The Revised Speech Perception in Noise Test (R-SPIN) in a Multiple Signal-to-Noise Ratio Paradigm. *Journal of the American Academy of Audiology*, 23(8), 590–605. <https://doi.org/10.3766/jaaa.23.7.9>
- Wright, J., Ciccone, D. S., & Brelsford, J. W. (1977). Selective encoding processes in recognition. *British Journal of Psychology*, 68(3), 289–295. <https://doi.org/10.1111/j.2044-8295.1977.tb01592.x>
- Yasmin, S., Irisk, V. C., Johnsrude, I. S., & Herrmann, B. (2023). The effects of speech masking on neural tracking of acoustic and semantic features of natural speech. *Neuropsychologia*, 186, Article 108584. <https://doi.org/10.1016/j.neuropsychologia.2023.108584>
- Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction Error Associated with the Perceptual Segmentation of Naturalistic Events. *Journal of Cognitive Neuroscience*, 23(12), 4057–4066. https://doi.org/10.1162/jocn_a.00078
- Zacks, J. M., Speer, N. K., & 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., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and Aging*, 21(3), 466–482. <https://doi.org/10.1037/0882-7974.21.3.466>
- Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, 127(1), 3–21. <https://doi.org/10.1037/0033-2909.127.1.3>
- Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, 130(1), 29–58. <https://doi.org/10.1037/0096-3445.130.1.29>
- Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive Load During Speech Perception in Noise: The Influence of Age, Hearing Loss, and Cognition on the Pupil Response. *Ear and Hearing*, 32(4), 498–510. <https://doi.org/10.1097/AUD.0b013e31820512bb>
- Zellner, B. (1994). Pauses and the Temporal Structure of Speech. In *Fundamentals of speech synthesis and speech recognition*. Wiley & sons.
- Zhang, Q., Griffiths, T. L., & Norman, K. A. (2023). Optimal policies for free recall. *Psychological Review*, 130(4), 1104–1124. <https://doi.org/10.1037/rev0000375>
- Zwaan, R. A. (1996). Processing Narrative Time Shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1196–1207. <https://doi.org/10.1037/0278-7393.22.5.1196>
- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), 386–397. <https://doi.org/10.1037/0278-7393.21.2.386>
- 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>