Acoustic features drive event segmentation in speech

Omri Raccah, Keith Doelling, Lila Davachi, David Poeppel

▶ **To cite this version:**

Omri Raccah, Keith Doelling, Lila Davachi, David Poeppel. Acoustic features drive event segmentation in speech. Journal of Experimental Psychology: Learning, Memory, and Cognition, 2022, 49 (9), pp.1494-1504. 10.1037/xlm0001150 . pasteur-03924690

HAL Id: pasteur-03924690

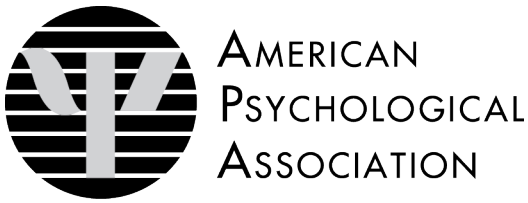
https://pasteur.hal.science/pasteur-03924690v1

Submitted on 5 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



Journal of Experimental Psychology: Learning, Memory, and Cognition

Manuscript version of

Acoustic Features Drive Event Segmentation in Speech

Omri Raccah, Keith B. Doelling, Lila Davachi, David Poeppel

Funded by:

- Fyssen Foundation
- National Institutes of Health
- National Science Foundation

© 2022, American Psychological Association. This manuscript is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final version of record is available via its DOI: https://dx.doi.org/10.1037/xlm0001150

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Acoustic features drive event segmentation in speech

Omri Raccah[1*], Keith Doelling[2,3], Lila Davachi[5,6], David Poeppel[1,3,4]

[1]Department of Psychology, New York University, New York, NY, USA

[2]Institut Pasteur, Institut de l'Audition, Paris, France

[3]Center for Language, Music, and Emotion, NYU & Max Planck Institute

[4]Ernst Strüngmann Institute (ESI) for Neuroscience, Frankfurt, Germany

[5]Department of Psychology, Columbia University, New York, NY, USA

[6]Nathan Kline Institute for Psychiatric Science, Orangeburg, NY, USA

*****Correspondence to**:** OR (or409@nyu.edu)

Author Contributions: O.R. and K.D. conceived of the study. D.P. and L.D. supervised the experiments and data analysis. O.R. and K.D. analyzed the data. O.R. wrote the paper. All authors contributed to the writing by reviewing and editing the manuscript.

Acknowledgments: National Science Foundation Graduate Research Fellowship to O.R. (DGE 1839302). Fyssen Foundation Postdoctoral Fellowship to K.D. We would like to thank Dr. Gwyneth Lewis for compiling the word stimuli used in this study. We also thank Niki Lam for assistance with data collection.

Conflict of interest: the authors declare no competing financial interests

Abstract

While our perceptual experience seems to unfold continuously over time, episodic memory preserves distinct events for storage and recollection. Previous work shows that stability in encoding context serves to temporally bind individual items into sequential composite events. Importantly, this phenomenon has been almost exclusively studied using visual and spatial memory paradigms. Here we adapt these paradigms to test the role of speaker regularity for event segmentation of complex auditory information. The results of our auditory paradigm replicate the findings in other sensory modalities – finding greater within-event temporal memory for items within speaker-bound events and greater source memory for items at speaker or event transitions. The task we employ significantly extends the ecological validity of past paradigms by allowing participants to encode the stimuli without any suggestions on the part of the experimenter. This unique property of our design reveals that, while memory performance is strongly dependent on self-reported mnemonic strategy, behavioral effects associated with event segmentation are robust to changes in mnemonic strategy. Finally, we consider the effect of serial position on segmentation effects during encoding and present a modeling approach to estimate the independent contribution of event segmentation. These findings provide several lines of evidence suggesting that contextual stability in perceptual features drives segmentation during word listening and supports a modality-independent role for mechanisms involved in event segmentation.

Introduction

It has long been shown that humans perceive and can identify boundaries between every day experiences (Kurby & Zacks, 2008; Newtson, 1973; Zacks et al., 2007; Zacks & Swallow, 2007). Over the past decade, research has uncovered that context shifts in stimulus properties or goal-state have marked consequence for the relationship between items in long-term memory (Clewett & Davachi, 2017; DuBrow & Davachi, 2016; Ezzyat & Davachi, 2011; Howard & Kahana, 2002; Polyn et al., 2009). Specifically, stability in environmental properties contribute to better cued and serial memory for items belonging to the same context compared to those spanning a contextual boundary (Clewett & Davachi, 2017; Clewett et al., 2020; DuBrow & Davachi, 2013, 2016; Heusser et al., 2018; Horner et al., 2016). Furthermore, enhanced associative memory has been observed for items occurring at context shifts or boundaries (Swallow et al., 2011; Swallow et al., 2009; Zwaan, 1996). Together, this work underscores that event structure can arise from context changes in the absence of a surprising stimulus or an explicit prediction error (DuBrow & Davachi, 2016).

Event segmentation has been almost exclusively investigated using visual and spatial memory paradigms (Baldassano et al., 2017; Clewett & Davachi, 2017; Horner et al., 2016; Howard & Kahana, 2002; Olman et al., 2009; Ranganath & Rainer, 2003). As in spatial navigation, complex auditory signals, such as spoken language and music, contain composite events with ordered constituents; however, little is known about the environmental conditions which facilitate event segmentation for such signals. Unlike visual and spatial domains, the auditory system relies on constant acoustic change for

effective perception. Furthermore, prior work has demonstrated marked differences in memory abilities across vision and audition (Cohen et al., 2011; Cohen et al., 2009; Morey & Mall, 2012; Xu et al., 2020), suggesting a possible asymmetry for encoding these signals. It is therefore unclear whether event segmentation in the auditory domain also relies on context shifts in perceptual properties. This domain of research generally defines event boundaries as shifts in perceptual context, and we follow this sense here. Similarly, we refer to the sequentially bounded representations which emerge from such boundaries as events or episodes in memory. In the current study, we investigate the role of transitions in speakers for segmenting individually spoken words into distinct episodes in memory. To this end, we adapt procedures from sequence memory paradigms (Clewett & Davachi, 2017; Heusser et al., 2018), which leverage *source memory* and *temporal order memory* performance as markers for event segmentation under controlled settings. Given the multisensory nature of events in everyday life, we expect that segmentation effects in visual memory paradigms should extend to auditory sequences under controlled conditions.

The current paradigm addresses limitations in the ecological validity of typical approaches by allowing participants to encode the stimuli without explicit task instructions or suggested strategy on the part of the experimenter. This divergence from previous research isolates the effect of perceptual context unconfounded by task-related changes across event boundaries (e.g. when participants are asked to provide subjective ratings with respect to stimulus features; (Clewett et al., 2020; DuBrow & Davachi, 2016; Heusser et al., 2018; Heusser et al., 2016; Pu et al., 2022; Sols et al., 2017; Wen & Egner, 2022). Furthermore, this approach allows us to examine how

unprompted mnemonic strategies influence both overall memory performance as well as event segmentation behavior. Notably, research into the role of reward prediction errors (as opposed to shifts in perceptual context) have shown significant event segmentation effects without explicit encoding instructions (Rouhani et al., 2018; Rouhani et al., 2020), and naturalistic studies into event segmentation often involve passive listening or viewing as well (Baldassano et al., 2017; Michelmann et al., 2021). Finally, we consider serial position effects (Howard & Kahana, 2002; Murdock, 1962) as a potential confounding factor in our paradigm and use a model fitting approach to estimate segmentation effects independent of serial position effects.

An account for event segmentation during spoken word listening serves to considerably advance our understanding of sequential representation and operations extensively studied in the auditory domain (Dehaene et al., 2015). Furthermore, such work expands on the modality-dependent nature of behavioral and neuroscientific findings from visual memory studies. To our knowledge, the consequences of such modality-dependent differences (if any) have not been systematically explored in the literature.

Materials and Methods

Participants

Native English-speaking participants (N = 56; 37 females; mean age 24.8 y, range 18-53) were recruited from New York University and the New York Metropolitan Area. In total, the experiment lasted approximately one hour. Subjects were compensated for their participation. The study was approved by the local institutional

review board (New York University's Committee on Activities Involving Human Subjects).

Auditory Stimuli

The materials included a set of 512 word stimuli, collected from the English Lexicon Project (ELP), which could be categorized into groups based on semantic and linguistic features (Balota et al., 2007). Based on ELP's part-of-speech codes and capitalization, the set of words was reduced to a list containing only common nouns. Nouns were then arranged into eight groups based on the number of syllables and their written frequency. Based on the Hyperspace Analogue to Language (HAL) frequency norms (Balota et al., 2007; Lund & Burgess, 1996), words were binned into two equal groups based on their written-frequency: low-frequency referring to words with a log frequency of 7 or lower and high-frequency to words with a log frequency of 7 or higher.

Furthermore, words were grouped by semantic category (e.g., *plants, mammals, birds*). Category labels came from Battig and Montague's (B&M) category norming data (Van Overschelde et al., 2004), which consists of groups of word exemplars provided by participants in response to verbal prompts of roughly 56 categories (e.g., frequent responses to "weather" includes *climate, blizzard, cyclone, sunshine, thunder,* as exemplars). Items not listed in B&M were hand-labeled by a trained psycholinguist. While we did not directly manipulate semantic category in the current paradigm, this approach ensures a wide variety of category members across the experiment. After forming category groupings based on these criteria, we chose to include only consonant initial words, since few complete categories (requiring eight items per group) included

vowel-initial words. Finally, word groups had to share placement of the primary stress on the same syllable (e.g., begonia and hibiscus have stress on the 2nd syllable). For groups that did not have enough items, we looked to other sources for suitable words, including Wikipedia lists and the University of South Florida word association norms (Nelson et al., 2004).

The resulting word stimuli were recorded by four speakers (2 female and 2 male) whose native language is American English. The words were recorded in mono at 48kHz in a soundproof audio booth. Recordings were then preprocessed using Adobe Audition CC 2018. This included using a high-pass filter with a threshold at 80Hz to reduce low-frequency noise. Using a 30s silent recording before each stimulus recording session, ambient noise was regressed from the recording. Next, we applied an automatic audio segmentation protocol to parse the word stimuli from the continuous data file. A quality assessment was then performed to ensure successful word segmentation and to remove leftover noise in the audio files (clicks, skips, etc.). Finally, the word stimuli were exported to .wav format and signal amplitude was normalized across stimuli and speakers. To meet the constraints of the current experiment (described in *Task Design*), we further pruned each eight-item group to include only six items per group, based on results from a lexical decision task provided by the ELP (Balota et al., 2007). ***Table 1*** shows the example grouping for six stimuli sets used in the current experiment.

Table 1. Example groupings of items based on semantic category, frequency, and syllables.

*Low Frequency				High Frequency			
1-syl	2-syl	3-syl	4-syl	1-syl	2-syl	3-syl	4-syl
Food	Tool	Flower	Profession	Mammal	Relative	Place	Event
quiche	grinder	poinsettia	neurologist	dog	mother	gallery	reservation

mousse	scraper	hibiscus	cartographer	horse	sister	residence	graduation
flan	beater	hyacinth	stenographer	bear	cousin	hospital	celebration
scone	strainer	peony	geneticist	cat	father	restaurant	recreation
bisque	peeler	petunia	pathologist	bull	daughter	cinema	meditation
curd	cleaver	gardenia	technologist	fox	brother	studio	consultation

*Low frequency refers to log frequency < 7.0, high frequency indicates > 7.0 log frequency.

Task Design and procedures

To evaluate the role of speaker transitions in event segmentation, we extend a previously validated visual sequence memory paradigm (Heusser et al., 2018); **Figure 1**). The task was constructed using the Psychophysics Toolbox (http://psychtoolbox.org/) running on an Apple Macintosh OSX operating system. Participants underwent 16 experimental blocks; each consisted of an encoding session, followed by an order memory test, and finally a source memory test. During the encoding session, subjects listened to a list of 24 words, in which every six consecutive words were read by a single speaker before transitioning to a new speaker for the next six items. As such, the speakers' voices defined four perceptually distinct event-sequences of six successive words per event. In order to promote perceptual transitions across event boundaries, male and female voices were alternated in their presentation, such that a male voice was always followed by a female voice or vice versa. To control for word properties at the block-level, word stimuli were blocked such that words contained the same number of syllables and were either high or low written-frequency. This resulted in two blocks (48 total words) per stimulus class (e.g., 1 syllable, high written-frequency words). Given these constraints, stimuli were randomized across subjects, and speaker order was varied within blocks. As such, category memberships of the word stimuli (Table 1) were randomized such that only speakers' voices demarcated event boundaries. Word stimuli were preceded by a 1s ITI period and

followed by a 3s silent period before the onset of the next word. A fixation cross was presented at the center of the screen throughout the encoding session. In contrast to prior work (DuBrow & Davachi, 2016; Heusser et al., 2018), subjects were only instructed to memorize the order of the words and were not given an explicit task during the encoding phase.

Following the encoding session, subjects were given an order memory test, in which two previously heard words were *visually* presented side by side. Participants were asked to indicate which word was presented first during the encoding phase and also indicate their memory confidence (high/low confidence, HC/LC). Hence, there were a total of four responses during each test trial (right first HC, right first LC, left first HC, or left first LC). Each temporal memory test consisted of four 'within-event' word pairs (one per event) and three 'across-event' word pairs. Specifically, the 2^{nd} and 6^{th} items were tested within-events, while across-event pairs consisted of the 5^{th} element in an event and the 3^{rd} element in the following event. As such, within- and across test pairs were always separated by three intervening word items. The visual order (from right to left) in which words were shown was randomized, while ensuring that half of all within as well as across test trials were presented in both orders across task blocks.

Immediately after the temporal order test, subjects were given a source memory test. Here, participants were shown a single, previously presented word and asked to indicate whether the word was read by a female or male speaker. Subjects were likewise asked to provide confidence judgements. The 1^{st} or 4^{th} items within each event were tested on each trial, representing a boundary and non-boundary condition, respectively. In both the temporal order and source memory tests, the initial half of the

items presented during encoding were tested first to reduce recency effects. In addition, test trials (unlike the encoding trials) were self-paced, with a .5s ITI between trials. Participants were given a practice block before starting the experiment, which was omitted from subsequent analyses.

Figure 1. Auditory event boundary task. (1) During each block (16 total), participants listened to a series of 24 words. Each set of six consecutive items was read by a distinct speaker (alternating male/female speakers). (2) This was followed by an order memory test to assess temporal memory for items within (2^{nd} and 6^{th}) and across event-boundaries (5^{th} and 3^{rd}). (3) Next, subjects were given a source memory test and asked to indicate whether a specific word was said by a male or female speaker. Source memory was tested for the 1^{st} (boundary) and 4^{th} (non-boundary) items within each event. Items in both the temporal order and source memory tests were shown visually.

Post-task questionnaire and unprompted mnemonic strategies

After participants completed the experiment, a questionnaire was administered to evaluate subjects' understanding of task as well as naturally-adopted mnemonic strategies. First, subjects were asked to provide an open response stating their strategy for memorizing word order, and whether this self-reported strategy changed throughout the task. The open-ended questions were (1) *What is the general strategy you used to memorize the order of the words?* and (2) *Did this strategy change throughout the task? If so, please specify approximately which block(s) your strategy changed.*

Participants were then instructed to provide responses to statements on a Likert scale, which included five balanced responses (*strongly disagree, disagree, neither agree nor disagree, or strongly agree*). The first two questions serve to assess subjects' understanding and subjective difficulty of the task: (1) *I understood the instructions of*

the task and (2) *I found the task difficult*. Four subsequent statements were included to evaluate the degree to which subjects used specific unprompted mnemonic strategies to memorize temporal order across item pairs. The statements were composed of strategies suggested for participants in previous experiments (Clewett et al., 2020; Ezzyat & Davachi, 2014; Heusser et al., 2018) and other common mnemonic approaches. In particular, the statements included four categorical designations:

i. Story-telling: *I created stories to memorize the order of the word,*
ii. Method of loci: *I used imagined spatial cues (or landmarks) to memorize the order of the words*,
iii. Associative binding: *I imagined the words interacting with each other to memorize their order*,
iv. Rehearsal: *I continuously repeated the words in my mind to memorize their order*.

Subjective reports for these four statements were analyzed with respect to overall memory performance and segmentation behavior (**Figure 5**).

Modeling of segmentation and serial position effects

We sought to understand to what extent serial position effects at the list-level (i.e. primacy/recency effects; (Howard & Kahana, 2002; Murdock, 1962)) and segmentation effects are both present in our data. To this end, we applied a model fitting approach to jointly estimate these two effects at the subject-level average data, considering mean performance for items in the order during encoding (see **Figure 3**). We modeled serial

position effects (primacy/recency; *PR Model*) as a 2nd order polynomial (convex for accuracy and concave for RT). For the PR Model, we introduce a nonlinearity so that the model follows import properties of the data: for accuracy, a sigmoid squashing function to limit the outputs to between 0 and 1, and for response times, an exponential function to fit typical RT distributions. We constructed an additional model which consists of a step function to estimate event segmentation effects across the sequential items in a given encoding period (*ES Model*). As such, the step function captures memory accuracy and response time for order memory (within versus across) or source memory (boundary versus non-boundary). Importantly, while we fit the PR Model independently to the data, we additionally exploit a model which linearly combines these two model fits (*Combined Model*):

$$ES(x) = s\mathcal{X}_A(x) \qquad \mathcal{X}_A(x) = \begin{cases} 1 & if \quad x \nmid 2 \\ -1 & if \quad x \mid 2 \end{cases}$$

$$PR(x) = a_1x^2 + a_2x + c$$

$$CM(x) = ES(x) + PR(x)$$

$\mathcal{X}_A(x)$ alternates in its conditions depending on whether the item is in the odd or even position (corresponding to within or across comparison). Importantly, the predicted effects of segmentation are flipped for RTs compared to accuracy data given opposite patterns of expected segmentation effects (i.e. faster RTs and greater accuracy within-events). This can be represented in the model by a flip of the sign of the step parameter, s. We fit the parameters to subject-level data by minimizing mean squared error of each function using the Nelder-Mead algorithm (Gao & Han, 2012). After which, we tested the independent role of event segmentation in two complementary analyses.

First, we estimated the parameter fit for the step function within the Combined Model and tested significance for this parameter relative to a null distribution. Second, we fit the PR Model and subtract this model fit from the raw subject-level data, reducing the influence of primacy/recency effects. Subsequently, we recomputed our effects of interest on the residual data (within vs. across).

Statistical testing

Throughout, statistical analysis was applied using nonparametric permutation tests. We used paired permutation tests when computing group-level results given our within-subject design. Across these comparisons, we implemented 10,000 permutations to ensure a reliable estimation of the null distribution. Significance was evaluated at $p < 0.05$. Note, we indicate the use of a one-tailed test when an effect is evaluated in a specific direction, otherwise a two-tailed statistic is reported.

Results

Effects of speaker-bound events on temporal order and source memory

We first tested whether speaker event boundaries drive event segmentation, as approximated by temporal order and source memory performance. In keeping with findings in visual memory studies (Heusser et al., 2018), we hypothesized greater within-event temporal memory for items within contextually bounded events and greater source memory for items at event boundaries. While the subsequent analysis considered binned data across stimulus features (syllable-length and written-frequency)

and subjective confidence, we report several notable effects on memory performance, which are described in detail in the Supplementary Section (**Figure S1**; **Figure S2**).

Consistent with the hypothesis, we found higher order memory performance for items belonging to the same event compared to those across speaker boundaries ($t(56) = 3.89$, $p < 0.001$) (**Figure 2A**). In addition, we showed that subjects' response times during retrieval (**Figure 2B**) were significantly slower when recalling serial order across speaker event boundaries ($t(56) = -4.65$, $p < 0.001$). These findings indicate that perceptual context directly modulates order memory performance, such that items studied within the same speaker context were better remembered and retrieved more quickly. For source memory, also consistent with prior work (Clewett & Davachi, 2017; Heusser et al., 2018; Speer & Zacks, 2005), we found that accuracy was significantly higher for boundary compared with non-boundary items (**Figure 2A**; $t(56) = 6.92$, $p < 0.001$). Further, subjects' response times (**Figure 2B**) were slower for non-boundary than boundary item source attribution ($t(56) = -4.84$; $p < 0.001$). Notably, slower reaction time during source memory could be related to differences in looking times at event boundaries (Hard et al., 2011). Finally, using a two-way repeated measures analysis of variance (rmANOVA), we found a significant task (source/order) by condition (boundary/across and nonboundary/within) interaction for memory accuracy ($F(56) = 46.32$, $p < 0.001$) and response time ($F(56) = 31.44$, $p < 0.001$). Together, these data suggest that speaker transitions serve to diminish temporal order performance across these boundaries while concurrently improving source attribution at event boundaries, indicating a potential trade-off between these memory processes. This finding from audition provides direct quantitative confirmation of event segmentation effects

identified in visual encoding paradigms, which construct perceptual events through embedding images in colored frames or through employing image categories (e.g. objects and faces) (Heusser et al., 2018; Sols et al., 2017; Wen & Egner, 2022).

Figure 2. Group-level temporal and source memory findings. We find significantly higher order memory performance (proportion correct) and faster overall RT for items within speaker-bound events relative to across events. Source memory performance for boundary items was significantly higher compared to non-boundary items. Boundary items additionally show significantly faster mean response times. A significant task (source/order) by condition (boundary/across and nonboundary/within) interaction was found for both percent correct and response time. Error bars denote 95% confidence intervals. ** P < .001.

Segmentation effects through the lens of serial position

A critical open question concerns how temporal order memory and source memory effects are modulated as a function of serial position during encoding (**Figure 3A**). To our knowledge, the dynamic variance in event segmentation effects at the list level has not yet been reported in the literature. To estimate how memory effects are modulated as a function of serial position, we subdivided our data based on the test item position in each encoding list. For source memory, we considered accuracy and response time for boundary/nonboundary test items in the sequence these items were encoded for each block (**Figure 3B**). Similarly, for temporal order memory, we consider within- and across-event test items with respect to their serial position at the encoding list level **(Figure 3C**).

This approach revealed several data patterns which are obscured when considering the average data across experimental blocks (cf. **Figure 2**). In particular,

we find that, for both source and temporal order memory, segmentation effects emerge across adjacent test items during encoding (**Figure 3B** and **Figure 3C;** one-tailed test). In particular, we find that improved source attribution accuracy for boundary items is well captured across test stimuli belonging to the same event during encoding (event 1, B vs NB: $t(56) = 3.89$, $p < 0.001$; event 2, B vs NB: $t(56) = 1.63$, $p = 0.049$; event 3, B vs NB: $t(56) = 4.47$, $p < 0.001$; event 4, B vs NB: $t(56) = 3.01$, $p = 0.001$). Similarly, faster mean response times for boundary versus nonboundary items showed significant effects for items belonging to the same speaker-bound event (event 1, B vs NB: $t(56) = -2.15$, $p = 0.017$; event 2, B vs NB: $t(56) = -3.86$, $p < 0.001$; event 3, B vs NB: $t(56) = -2.05$, $p = 0.019$; event 4, B vs NB: $t(56) = -2.85$, $p = 0.002$). As such, we find that source attribution performance is modulated at the boundary (i.e. at the transition from one speaker to the next) across individual events.

In the case of temporal order memory, a somewhat different pattern emerges: we find that the first and final within-event comparisons during encoding appear to be modulated by serial position (**Figure 3C**). Specifically, accuracy for these within-event comparisons is significantly greater than for mid-block positions (i.e. positions 2 and 3; **Figure 3C**) (*event 1 within-event* vs. mid-block within- and across-comparisons, $p < 0.01$; *event 4 within-event* vs. mid-block within- and across-comparisons, $p < 0.01$). When considering only mid-block items, accuracy for within-events items were numerically higher overall but the pairwise comparisons showed a nonsignificant effect relative to across-event comparisons (*within position 2* vs. *across pos. 2*: $t(56) = 1.28$, $p = 0.1$; *within pos. 3* vs. *across pos. 3*: $t(56) = 0.65$, $p = 0.255$). Importantly, we did not find this pattern for across-event comparisons, possibly due to their serial position

during encoding, which never occupied the first or last items during encoding session (**Figure 3A**). Similarly, mean response time during serial order recall appeared to show a recency effect, with the final within-event comparison (event 4) showing significantly faster response times than all other within- and across-event comparisons (*event 4 within-event* vs. mid-block within- and across-comparisons, $p < 0.01$). Nevertheless, similar to source memory, we find that neighboring items during encoding indeed show faster response times for within- relative to across-event comparisons (*within pos. 1* vs. *across pos. 1*: $t(56) = -3.35$, $p < 0.001$; *within pos. 2* vs. *across pos. 2*: $t(56) = -2.95$, $p = 0.002$; *within pos. 3* vs. *across pos. 3*: $t(56) = -0.12$, $p = 0.454$). The effect found here for temporal order memory resembles primacy-recency effects widely reported in recognition and free-recall tasks (Howard & Kahana, 2002; Murdock, 1962). In the subsequent analysis, we aim to control for primacy-recency effects in order to estimate the individual contribution of event segmentation in our data.

Figure 3. Segmentation behavior is observed across adjacent encoding comparisons and is modulated by serial position. ***A***, encoding task structure in a single block. Four events were constructed across 24 four words. As described previously (see *Methods and Materials* & Figure 1), order memory was tested within (2^{nd} and 6^{th} items) and across event-boundaries (5^{th} and 3^{rd} items). Source memory was tested for the 1^{st} (boundary) and 4^{th} (non-boundary) items within each event. ***B***, source memory performance as a function of serial position during encoding. Mean accuracy and response time are captured for boundary/nonboundary items belonging to the same event. ***C***, temporal order memory as a function of serial position during encoding. Broadly speaking, serial order performance shows segmentation effects across adjacent test items; in this case, however, we additionally observe primacy/recency effects widely reported in recognition and free-recall paradigms (Howard & Kahana, 2002; Murdock, 1962). *Error bars denote 95% confidence intervals. * $P < 0.05$. ** $P < .001$.*

Model fitting reveals that serial effects interact with segmentation effects

The apparent primacy-recency effect during order memory retrieval raises a critical question: to what extent (if at all) are serial position effects and segmentation effects concurrently present in our data? We fit a model to our data to shed light on this question. Specifically, we fit two models: (1) a second-order polynomial function to capture the primacy-recency effects (PR Model) and (2) a step function to capture segmentation effects across adjacent comparisons (ES Model). We additionally linearly combine these two functions to construct a model which captures both effects (Combined Model).

After fitting these models to subject-level data (**Figure 4A** shows exemplar fit; **Figure S3** shows fit for each subject in our cohort), we apply two complementary analysis approaches to estimate the contribution of segmentation while controlling for

primacy-recency effects (see *Materials and Methods*). First, we fit the Combined Model for each memory type (order/source) and behavioral measure (accuracy/RT) (**Figure 4B**) and extract the step parameter s from the ES portion for each participant. As such, we are able to capture the contribution of the event segmentation parameter independently of the primacy-recency effects. We find that order memory accuracy and RT are significantly captured by the step parameter in the combined model (PC: param. = 0.02, $p = 0.0245$; RT: param. = -0.15, $p < 0.001$). Similarly, we find that accuracy and RT data during source retrieval are well captured by the step parameter (PC: param. = 0.04, $p < 0.001$; RT: param. = -0.13, $p < 0.001$).

In a complementary analysis, we fit the PR Model to individual subject data, which affords us a parameter estimate for the observed primacy-recency effect. Next, we subtract this model fit from each subjects' order memory accuracy and response time data and recompute the average subject comparison (within versus across) using the residual data, regressing out the primacy-recency effect computed for each individual subject. This approach revealed that residual accuracy is significantly greater for within than across comparisons ($t(56) = 2.29$, $p = 0.028$; **Figure 4C**). Furthermore, within comparisons showed significantly faster residual response time than across comparisons ($t(56) = -3.71$, $p < 0.001$; **Figure 4C**). Together, these complementary analyses provide evidence that segmentation effects are present in our data, even when controlling primacy-recency effects.

Figure 4. Modeling of serial position and segmentation effects. A, *combined (step + polynomial)* model fits for two exemplar subjects (see Figure S3 for model fit for all participants). Specifically, we fit a polynomial model to capture serial position as well as a combined model which additionally includes a step parameter to capture segmentation behavior (within versus across; boundary versus nonboundary). ***B,*** step (segmentation) parameter for combined model as compared to a null distribution for each memory type and behavioral measure. We find that the step parameter significantly captures our data independently of serial position effects. Gray dots indicate step parameter estimates for individual subjects and dashed lines represent group-level means. ***C,*** in a complementary analysis, we subtract the polynomial (or serial position) model from subject-level data. We find that this residual data preserves expected segmentation effects during serial order memory retrieval (i.e. within versus across; residual percent correct and RT). ** P < 0.05. ** P < .001.*

The effect of mnemonic strategy on memory performance and segmentation

In the current study, we sought to understand the role of naturally-adopted mnemonic strategies in source attribution and temporal order memory. Specifically, we tested how such strategies during encoding affect both overall memory accuracy and segmentation behavior. We evaluated participants' subjective reports of mnemonic strategy use with a post-task questionnaire (see *Materials and Methods*) for four key strategies. Three of the strategies involved mental imagery: story-telling, imagined spatial cues, and associative binding (imagining neighboring items interacting). The fourth strategy involved rehearsal through silent repetition.

We first tested the effect of mnemonic strategy on overall memory performance (using a Spearman's rank correlation). Overall memory performance was computed by binning trials across conditions for temporal order and source memory, respectively. We found that subjective rating of using a story-telling strategy best predicted overall

temporal memory performance (i.e., grouping within- and across-event conditions; Rho = 0.47, p < 0.001). We also found a robust correlation between reports of using imagined spatial cues and order memory performance (Rho = 0.42, p = 0.001). Rehearsal, showed a significant negative relationship with temporal memory accuracy (Rho =- 0.33, P = 0.015). Further, we found a trending but non-significant positive relationship between an associative binding strategy and memory performance (Rho = 0.26, p = 0.052). Overall source memory accuracy closely tracked the effect of mnemonic strategy as temporal order memory (*stories*: Rho = 0.31, p = 0.02; *spaces*: Rho = 0.38, p = 0.006; *associative binding*: Rho = 0.09, P = 0.52), with the exception of rehearsal, which showed a negative, but not significant effect (Rho = -0.23, p = 0.09). These findings suggest that naturally-adopted mnemonic strategies had specific and robust effects on subjects' overall accuracy which were consistent across memory types (**Figure 5**). To evaluate whether some strategies are employed together or in an opposing manner, we additionally performed a Spearman's correlation between individual strategies (P > 0.05; **Figure S4**). We found that adopting a story-telling strategy is significantly correlated with an associative binding strategy (Rho =0.55, P < 0.001), while other strategies did not show a significant correlation with one another (P > 0.05).

Next, we tested effect of mnemonic strategy on our behavioral markers for event-segmentation (DuBrow & Davachi, 2016). For each participant, segmentation was calculated as the mean difference between within- and across-event for the order memory test ($within - across$) and, for source memory, as the mean difference between the boundary and non-boundary conditions ($boundary - nonboundary$). As

such, greater positive values for these measures indicate a higher segmentation effect, or segmentation strength. Strikingly, we found that, while overall performance is robustly modulated by mnemonic strategy, segmentation strength shows no significant relationship across mnemonic strategies for temporal order (story-telling: Rho = -0.061, P = 0.66; method-of-loci: Rho = -0.02, P = 0.88; associative binding: Rho = -0.08, P = 0.54; rehearsal: Rho = 0.07, P = 0.64) or source memory (story-telling: Rho =-0.02, P =0.88; method-of-loci: Rho =-0.05, P =0.71; associative binding: Rho =-0.18, P =0.19; rehearsal: Rho =-0.08, P =0.54) (**Figure 5**).

Figure 5. Memory performance and segmentation strength as a function of naturally adopted mnemonic strategy. *We find that subjective reports of story-telling and spatial navigation strategies best predict source and temporal order memory performance. In contrast, rehearsal strategies show a negative trending relationship with overall source and temporal order performance. While self-reported mnemonic strategy appears to have a robust effect on overall memory performance, we find no relationship between strategies and segmentation strength. Shaded regions indicate 95% confidence intervals.*

Discussion

The present results show that memory effects of event segmentation are not limited to the visual domain (Clewett & Davachi, 2017; DuBrow & Davachi, 2016; Heusser et al., 2018) but extend to other modalities (**Figure 1**). Furthermore, we provide primary evidence that segmentation effects are driven by contextual stability in *perceptual* features as opposed to changes in internal state or decision criteria across event boundaries; these competing interpretations are inextricable in prior work, which

incorporates an encoding task in order to drive performance (e.g. pleasantness ratings, (DuBrow & Davachi, 2016; Heusser et al., 2018)). Furthermore, whereas in previous experiments participants were explicitly asked to adopt an associative binding strategy (Clewett et al., 2020; DuBrow & Davachi, 2016; Heusser et al., 2018), here we allowed them to adopt any mnemonic strategy (or none at all) as they saw fit. While the importance of goal-state and prediction in event segmentation is well-established in past literature (Antony et al., 2021; Ben-Yakov et al., 2021; Reynolds et al., 2007; Rouhani et al., 2019; Zwaan et al., 1995), the fact that the memory boost for within-event and boundary comparisons persisted despite these changes in task design suggests that event segmentation might be a more automatic process than previously argued, at least in the auditory domain. In other words, the lack of an explicit integration task suggest that participants may not need to engage in a conscious binding process for segmentation effects to arise. That these effects were not related to the strategies naturally adopted by our participants further emphasizes this point. Notably, however, the current paradigm does not rule out that participants' knowledge of an upcoming word-speaker source memory test drives event segmentation in an internal manner during encoding. That is, attention to the source (gender) to answer the source questions may promote the binding of presented words with their respective context over time, increasing segmentation at speaker transitions.

We additionally provide evidence that segmentation effects, at least in this design where there are no explicit requirements to change encoding at boundary items, are modulated by list-level serial position during encoding (**Figure 3**). In particular, our findings suggest that segmentation effects are captured across neighboring test items

during encoding (e.g. boundary and nonboundary items belonging to the same speaker-bound event). For temporal memory, we also observed list-level primacy/recency effects, generally reported in free-recall and recognition memory paradigms (Howard & Kahana, 2002; Murdock, 1962). Importantly, these effects were present for within-event comparisons but not for across-event memory test comparisons. Therefore, the primacy/recency effect observed for temporal order memory presents a potential confounding factor in our paradigm. We used a model fitting approach to address this concern. By fitting the data to a model of primacy and recency effects (PR model; 2^{nd} order polynomial) and a combined model incorporating an event segmentation model (ES model; step function), we were able to show that even when accounting for the effects of primacy and recency, the residual data still show a significant effect of event segmentation. We are therefore confident that the event-segmentation results are not due to a potential confound of primacy and recency. The present results can additionally inform computational models of event segmentation which rely on findings from free-recall paradigms to account for serial position effects (Rouhani et al., 2020).

We used a post-task questionnaire together with our passive encoding paradigm to test the effect of naturally-adopted mnemonic strategies on overall memory performance as well as event segmentation behavior. An analysis of this data revealed that story-telling and spatial memory strategies strongly predicted overall memory performance (both temporal and source memory), whereas rehearsal showed a negative relationship with overall serial order memory performance. The present results raise the question of why a rehearsal strategy displays a significant negative relationship with serial memory performance. Notably, we found no significant

relationship between adopting a rehearsal strategy and any other strategies (**Figure S4**), indicating that rehearsal is not simply opposing another more effective strategy (e.g. story-telling). Future work could investigate whether some trade-off in encoding – e.g. rehearsal enhancing primacy effects (Modigliani & Hedges, 1987; Reynolds & Houston, 1964) – is occurring with respect to serial memory accuracy. That said, we did not find a significant correlation between primacy strength – defined as the difference between within-event order accuracy for the 1st item pair and the subsequent (2nd) within-event item pair in each encoding block - and participants ratings for employing a rehearsal strategy (Rho = 0.075, P = 0.6).

Unlike overall memory performance, we find no effect of mnemonic strategy on behavioral markers of event segmentation (within-event vs. across-event and boundary vs. nonboundary). Together, these findings indicate that temporal and source memory accuracy is contingent on self-reported mnemonic strategy, whereas segmentation effects are seemingly robust to mnemonic strategy; this suggests that event segmentation may be largely independent of internal mnemonic strategies.

The present findings raise several critical questions regarding the underlying neural mechanisms that serve to segment ongoing perceptual experience in memory. Provided behavioral effects of segmentation can be reliably observed across sensory domains, the current work provides opportunities for bridging findings with previously reported neural signatures of event segmentation and retrieval (Baldassano et al., 2017; Clewett & Davachi, 2017; Hasselmo & Eichenbaum, 2005; Schapiro et al., 2014). In-depth efforts of this kind would help to further inform the modality-independent neural mechanisms that govern event segmentation and possibly how these systems

transform diverse perceptual signals to drive segmentation. Finally, our findings set the stage for further investigating event segmentation for complex auditory signals, such as hierarchically-organized language and music sequences (Dehaene et al., 2015; Hartley & Poeppel, 2020; Lerdahl & Jackendoff, 1996).

References

Antony, J. W., Hartshorne, T. H., Pomeroy, K., Gureckis, T. M., Hasson, U., McDougle, S. D., & Norman, K. A. (2021). Behavioral, physiological, and neural signatures of surprise during naturalistic sports viewing. *Neuron, 109*(2), 377-390. e377.

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. e705.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior research methods, 39*(3), 445-459.

Ben-Yakov, A., Smith, V., & Henson, R. (2021). The limited reach of surprise: Evidence against effects of surprise on memory for preceding elements of an event. *Psychonomic bulletin & review*, 1-12.

Clewett, D., & Davachi, L. (2017). The ebb and flow of experience determines the temporal structure of memory. *Current opinion in behavioral sciences, 17*, 186-193.

Clewett, D., Gasser, C., & Davachi, L. (2020). Pupil-linked arousal signals track the temporal organization of events in memory. *Nature communications, 11*(1), 1-14.

Cohen, M. A., Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). Auditory and visual memory in musicians and nonmusicians. *Psychonomic bulletin & review, 18*(3), 586-591.

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences, 106*(14), 6008-6010.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron, 88*(1), 2-19.

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

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

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological science, 22*(2), 243-252.

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

Gao, F., & Han, L. (2012). Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications, 51*(1), 259-277.

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General, 140*(4), 586.

Hartley, C. A., & Poeppel, D. (2020). Beyond the Stimulus: A Neurohumanities Approach to Language, Music, and Emotion. *Neuron, 108*(4), 597-599.

Hasselmo, M. E., & Eichenbaum, H. (2005). Hippocampal mechanisms for the context-dependent retrieval of episodes. *Neural networks, 18*(9), 1172-1190.

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

Heusser, A. C., Poeppel, D., Ezzyat, Y., & Davachi, L. (2016). Episodic sequence memory is supported by a theta–gamma phase code. *Nature neuroscience, 19*(10), 1374.

Horner, A. J., Bisby, J. A., Wang, A., Bogus, K., & Burgess, N. (2016). The role of spatial boundaries in shaping long-term event representations. *Cognition, 154*, 151-164.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46*(3), 269-299.

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

Lerdahl, F., & Jackendoff, R. S. (1996). *A generative theory of tonal music*. MIT press.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior research methods, instruments, & computers, 28*(2), 203-208.

Michelmann, S., Price, A. R., Aubrey, B., Strauss, C. K., Doyle, W. K., Friedman, D., Dugan, P. C., Devinsky, O., Devore, S., & Flinker, A. (2021). Moment-by-moment tracking of naturalistic learning and its underlying hippocampo-cortical interactions. *Nature communications, 12*(1), 1-15.

Modigliani, V., & Hedges, D. G. (1987). Distributed rehearsals and the primacy effect in single-trial free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(3), 426.

Morey, C. C., & Mall, J. T. (2012). Cross-domain interference costs during concurrent verbal and spatial serial memory tasks are asymmetric. *Quarterly Journal of Experimental Psychology, 65*(9), 1777-1797.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of experimental psychology, 64*(5), 482.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior research methods, instruments, & computers, 36*(3), 402-407.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology, 28*(1), 28.

Olman, C. A., Davachi, L., & Inati, S. (2009). Distortion and signal loss in medial temporal lobe. *PLoS One, 4*(12), e8160.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological review, 116*(1), 129.

Pu, Y., Kong, X.-Z., Ranganath, C., & Melloni, L. (2022). Event boundaries shape temporal organization of memory by resetting temporal context. *Nature communications, 13*(1), 1-13.

Ranganath, C., & Rainer, G. (2003). Cognitive neuroscience: Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience, 4*(3), 193.

Reynolds, J. H., & Houston, J. P. (1964). Rehearsal strategies and the primacy effect in serial learning. *Psychonomic Science, 1*(1), 279-280.

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.

Rouhani, N., Norman, K. A., & Niv, Y. (2018). Dissociable effects of surprising rewards on learning and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(9), 1430.

Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2019). Reward prediction errors create event boundaries in memory. *bioRxiv*, 725440.

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

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of cognitive neuroscience, 26*(8), 1736-1747.

Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. (2017). Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. *Current Biology, 27*(22), 3499-3504. e3494.

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.

Swallow, K. M., Barch, D. M., Head, D., Maley, C. J., Holder, D., & Zacks, J. M. (2011). Changes in events alter how people remember recent information. *Journal of cognitive neuroscience, 23*(5), 1052-1064.

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.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language, 50*(3), 289-335.

Wen, T., & Egner, T. (2022). Retrieval context determines whether event boundaries impair or enhance temporal order memory. *bioRxiv*, 2022.2001.2002.474709. https://doi.org/10.1101/2022.01.02.474709

Xu, M., Fu, Y., Yu, J., Zhu, P., Shen, M., & Chen, H. (2020). Source information is inherently linked to working memory representation for auditory but not for visual stimuli. *Cognition, 197*, 104160.

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.

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current directions in psychological science, 16*(2), 80-84.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1196.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological science, 6*(5), 292-297.