

Do actions structure auditory memory? Action-based event segmentation effects on sensory responses, pupil dilation and sequential memory

Marta Font-Alaminos¹, Nadia Paraskevoudi¹, Jordi Costa-Faidella¹, and Iria SanMiguel¹

¹Brainlab-Cognitive Neuroscience Research Group

July 14, 2023

Abstract

Our actions shape our everyday experience: what we experience, how we perceive and remember it, is deeply affected by how we interact with the world. Performing an action to deliver a stimulus engages neurophysiological processes which are reflected in the modulation of sensory and pupil responses. In this study, we hypothesized that these processes shape memory encoding, parsing the experience by grouping self- and externally-generated stimuli into differentiated events. Participants encoded sound sequences, in which either the first or last few sounds were self-generated and the rest externally-generated. We tested recall of the sequential order of sounds that had originated from the same (within event) or different sources (across events). Memory performance was not higher for within event sounds, suggesting that the memory representation was not structured by actions. However, during encoding, we replicated the well-known electrophysiological response attenuation, together with increased pupil dilation for self-generated sounds. Moreover, we found that at the boundary between events, physiological responses to the first sound originating from the new source were determined by the direction of the source switch. The results suggest that introducing actions, acts as a stronger contextual shift than removing them, despite not directly contributing to memory performance. The findings contribute to our understanding of how interacting with sensory input shapes our experiences, by addressing the unexplored relationships between action effects on sensory responses, pupil dilation and memory encoding, and discarding a meaningful contribution of low-level neurophysiological mechanisms associated to action execution in the modulation of memory.

1. Introduction

The way we experience the world is fundamentally shaped by our actions. Numerous studies have demonstrated that self-generated (SG) stimuli elicit reduced subjective sensations and sensory responses compared to stimuli that are externally-generated (EG; for reviews see Hughes et al., 2013; Schröger et al., 2015). In the auditory domain, the effects of self-generation on sensory responses have been mainly studied on event-related potentials (ERPs), where an attenuation of the N1 component is consistently observed for SG sounds. Although the self-generation effect has been mostly attributed to motor-driven predictive processes engaged via forward modelling (Blakemore et al., 1998; Frith et al., 2000; Wolpert et al., 1995), this modulation also occurs in nonpredictive situations (e.g. Horváth et al., 2012; Paraskevoudi & SanMiguel, 2023a) indicating that other mechanisms could be mediating the sensory attenuation. Recently, Paraskevoudi & SanMiguel (2023a, 2023b) suggested that subcortical neuromodulation during motor actions (Aston-Jones & Cohen, 2005; Eggermann et al., 2014; McGinley et al., 2015; Vinck et al., 2015) may regulate the sensory attenuation observed during actions. This hypothesis originated from findings demonstrating a strong correlation between pupil diameter, an indicator of locus coeruleus norepinephrine system (LC-NE) activity, and actions (Lubinus et al., 2022; Yebra et al., 2019). Although Paraskevoudi & SanMiguel (2023a, 2023b) provided preliminary evidence linking self-generation effects and pupil dilation during sound processing, further

research is required to identify the specific contribution of LC-NE activity, proxied by pupil dilation, to the self-generation effect.

Despite the elusive neural underpinnings of the self-generation effect, existing research provides compelling evidence to suggest that active production not only influences perception, but also memory (McCurdy et al., 2017). For instance, played piano melodies and spoken words are better remembered than the same stimuli when they are encoded in the absence of overt actions, a phenomenon known as the production effect (MacDonald & MacLeod, 1998; MacLeod, 2011; MacLeod & Bodner, 2017). The production effect is frequently explained with the distinctiveness model, which proposes that items containing more distinct components have a higher probability of being recalled from memory than items with fewer components (Ozubko & Macleod, 2010). Once again, a recent study has linked memory enhancements resulting from overt actions to the activation of the LC-NE system, reflected in increased pupil dilation (Yebra et al., 2019). All in all, it is possible that the distinctiveness provided by the actions, along with the differential sensory encoding of SG sounds, could be contributing to the memory performance enhancement seen in the production effect. However, to date, only three studies have combined the typical paradigms employed to measure self-generation effects with a memory task in order to test the relationship between the self-generation effects (on sensory processing) and the production effects (on memory). Using slightly different memory tasks and action-sound relationships, these studies have found conflicting results ranging from no difference in memory (Font-Alaminos et al., 2023) to memory enhancement (Paraskevoudi & SanMiguel, 2023b) and memory impairment (Paraskevoudi & SanMiguel, 2023a) for SG compared to EG sounds, despite the presence of reliable self-generation effects on auditory ERPs in all three studies. Thus, further research is needed to clarify the impact of actions on sensory processing and memory encoding and identify the critical variables responsible for the discrepancy in the observed effects.

An intriguing possibility is that actions may not simply modulate the strength of memory traces to specific items, but perhaps they contribute to structure the memory storage. Memory, a crucial aspect of human cognition, enables individuals to store and retrieve past experiences that shape their present and future behaviors. These experiences are structured and organized in discrete units or events based on contextual cues, a process called event segmentation (DuBrow & Davachi, 2013). Event segmentation allows for the effective organization and retrieval of episodic memories by breaking them down into meaningful units or events (Zacks & Swallow, 2007). These meaningful units are separated by event boundaries, identifiable moments where there is a noticeable change in the content, context, or purpose of an experience (Kurby & Zacks, 2008). Minor changes in the physical environment or in the ongoing task, such as on the background color or sound, can create these event boundaries, highlighting their importance in the organization of episodic memory (Kurby & Zacks, 2008; Newtonson & Engquist, 1976; Speer et al., 2003; Swallow et al., 2010). However, to our knowledge, no studies have specifically investigated the potential of actions per se to create event boundaries. Considering the current self-generation-related research on sensory processing, neuromodulation, and memory it seems plausible to assume that actions may not only modulate the strength of memories for SG stimuli, but they may additionally structure the encoding of sounds in memory, possibly generating differentiated storages for SG and EG sounds.

The primary aim of this study was to examine the impact of motor actions on memory encoding. We hypothesize that neuromodulatory mechanisms tied to actions structure memory encoding and lead to the formation of distinct memory representations for SG and EG auditory stimuli. Previous studies have provided evidence for event segmentation during memory encoding demonstrating that, when asked to remember the sequential order of stimuli, items encoded as belonging to the same event present increased memory performance compared to those belonging to different events (Clewett et al., 2019, 2020; DuBrow & Davachi, 2016; Pettijohn et al., 2016; Radvansky & Zacks, 2017). Here, we hypothesize that differentiated events may be defined by the presence or absence of a motor act to deliver the stimuli. In order to test this hypothesis, we assessed temporal order memory performance for sound pairs encoded within the same event versus across separate events.

Furthermore, we extended on previous research by investigating whether SG sounds confer a memory ad-

vantage over EG sounds when delivered concurrently to form unified events. Previous studies attempting to relate self-generation and production effects have tested the effects of self-generation on memory recall of discrete sounds (Font-Alaminos et al., 2023; Paraskevoudi & SanMiguel, 2023a, 2023b). Here, we tested the effects on sequential memory, predicting that SG-SG pairs would elicit better temporal order memory performance than EG-EG pairs. To examine the effects of self-generation during encoding, we analyzed sensory ERP components and expected to replicate the well-established finding of attenuated sensory responses to SG sounds (Schröger et al., 2015). We also measured pupil diameter to investigate the neuromodulatory effects surrounding the motor actions and hypothesized that greater pupil dilation would be observed for SG stimuli compared to EG stimuli during sound encoding (Paraskevoudi & SanMiguel, 2023a, 2023b).

To explore the influence of actions on event segmentation, we examined whether a boundary was formed after a change in the sound’s source (EG or SG). In line with context change literature (Donchin & Coles, 1988; Polich, 2007), we predicted an enhanced P3 on the ERP and greater pupil dilation for the boundary sound compared to non-boundary sounds. Finally, we investigated the correlation between memory performance and the magnitude of the self-generation and pupil dilation effects, and explored whether a relationship exists between the effects on pupil and on electrophysiological sensory responses. This study was pre-registered as part of the Open Science Framework (<https://doi.org/10.17605/OSF.IO/5ZUGV>).

Methods

Participants

The sample size for this study was determined based on previous similar studies (Silva et al., 2019). Initially, a total of 28 healthy participants were recruited to take part in the experiment. However, 3 participants had to be excluded due to a low signal-to-noise ratio in the EEG data. As a result, the final sample for this study consisted of 25 participants (mean age: 21 years old; 8 males). All participants were compensated for their time with a payment of 10 euros per hour. Participants were mainly recruited among University of Barcelona students. The study excluded participants with hearing impairments, psychiatric or neurological disorders, age below 18 years or over 50 years, recent consumption of drugs or pharmaceuticals that affect the central nervous system, or inability to comply with the task instructions. In addition, participants who wore prescription glasses or contact lenses were excluded to ensure clean pupillometry data. The Bioethics Committee of the University of Barcelona (IRB00003099) approved the study, and all participants provided informed consent after completing a general health questionnaire.

Study Design

We designed an electrophysiological and pupillometry experiment with a temporal order memory task to investigate how self-generation affects event segmentation. The order memory task had three phases: encoding, retention, and retrieval (Figure 1). During the *encoding* phase, participants were presented with a sequence of nine sounds that included both SG and EG sounds. Critically, all SG or EG sounds at encoding appeared consecutively, creating two different events based on the source of the sounds within the nine-sound sequence. Immediately after, there was a short *retention* period to memorize the sequential order of the sounds. During the *retrieval* phase, participants were presented with a pair of sounds, and they had to indicate whether they appeared in the same order as during encoding. The test sounds belonged either to the same event (within condition) or to different events (across condition) but were non-consecutive in the sequence. We used visual stimuli to structure the task and indicate participants when to perform actions to generate sounds. To ensure participants understood the task, they underwent at least five practice trials using a simplified version of the task before the start of the experiment (see section 2.4.).

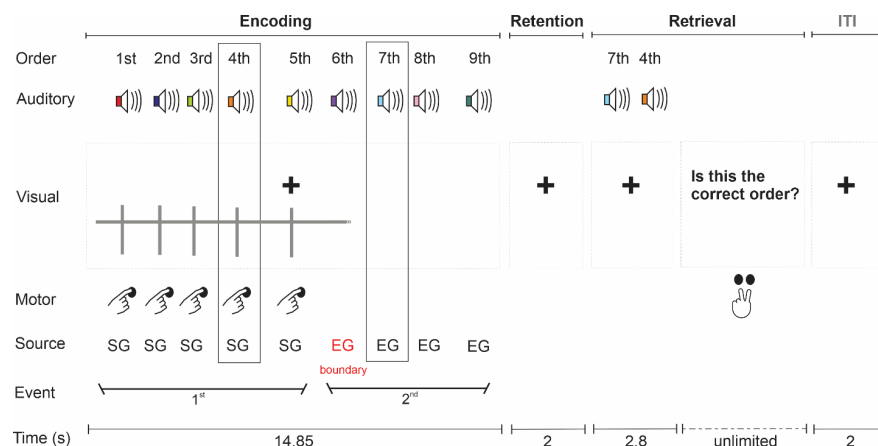


Figure 1. Schematic description of the temporal order memory task with externally-generated (EG) and self-generated (SG) sounds. This example trial corresponds to the across-events condition. The timepoints mark the duration of each phase of the trial. ITI corresponds to the inter-trial interval.

Encoding phase

At the beginning of each trial, participants were presented with a row of vertical lines on the screen. The number and distribution of vertical lines varied from trial to trial as determined by the experimental manipulations. Throughout the encoding period (14.85 s), a horizontal line moved steadily across the screen from left to right, passing over each vertical line in its path. Participants were instructed to press the bottom-right button on a response pad with their right index finger whenever the horizontal line crossed a vertical line. This action produced a sound immediately after the button press. In each trial, the encoding set consisted of nine sounds, some of which were delivered consecutively after each button press (SG), while others were presented passively (EG) either before or after the sequence of button presses. This manipulation created two events in a sequence, each event formed by SG or EG sounds. The change in the source that generated the auditory stimuli, either EG or SG, acted as the contextual shift, and the first sound delivered by the new source was the boundary sound. Participants were instructed to pay attention to the order of all the sounds regardless of the motor task. The sound onset asynchrony was set between 1.3 to 2 s semi randomly in steps of 0.05 s.

Retention and Retrieval phase

Following the encoding phase, participants were given a short retention period of 2 s to retain the sequence's order. During the retrieval phase, a pair of two nonconsecutive sounds of the encoding phase were presented, with a 2 s onset separation interval. The pair of test sounds could appear in the correct or in the inverse order of presentation and could both be from the same encoding event (*within* condition) or could be one sound from each of the two encoding events (*across* condition). After 0.8 s following the onset of the second test sound, participants were asked if the pair's order of presentation was the same as at encoding. They had unlimited time to respond on the response pad YES or NO using the middle and index fingers of their left hand. The identity of the two buttons, either YES or NO, was counterbalanced across participants. The inter-trial interval was set to 2 s after the participant's response.

Visual Stimuli

The use of visual stimulation was necessary to ensure participants followed instructions and maintained fixation during the experiment, which allowed us to record their pupil response accurately. We presented a fixation cross at the center of the screen throughout the experiment, to indicate where participants should focus their gaze. The moving horizontal line from left to right and the stationary vertical lines appeared just below the fixation cross to indicate the progress of each trial and the appropriate time to press the button.

These visual stimuli appeared within a small area on the center of the screen (visual angle 2.3°) to eliminate any effects of gaze position on pupil diameter (Gagl et al., 2011).

Auditory Stimuli

We developed a set of unique auditory stimuli for our experiment by sourcing identifiable sounds from several freely available sound databases (Adobe; FreeSound; Belin et al., 2000; Gygi & Shafiro, 2010; Hocking et al., 2013; Norman-Haignere et al., 2015). Initially, we extracted 1110 different sounds, which we edited to have a duration of 250 ms and 10 ms exponential onset/offset ramp. To ensure the task’s feasibility, we manually selected sounds that were still recognizable after editing (i.e., not noise) and categorized them into distinct semantic groups. We grouped similar sounds into the same category to create as many different categories of sounds as possible, resulting in 15 semantic categories with 24 sounds each from animal, environmental, human, and music sounds (e.g. high pitch animal vocalizations, drums, alarms, aspirated vowels, etc.). During practice trials, we presented participants with pure tones of different frequencies ranging from 300 to 3700 Hz in 100 Hz steps instead of experimental sounds. All sounds were played at a sampling rate of 44.1 kHz, 16 bits, mono, and an intensity of 75 db.

We generated 24 unique sound sequences for each participant, consisting of 9 different sounds randomly selected without replacement from 9 of the 15 semantic categories to ensure variability. These unique sound sequences were presented 9 times each with different experimental manipulations (see section 2.2.6.). In each unique sound sequence, seven of the sounds were presented only in the encoding phase (encoding-only sounds), while the remaining two encoding sounds were also test sounds, presented again during the retrieval phase. To prevent sequential presentation of the 9 repetitions for each unique sequence, we distributed the 216 trials into 24 blocks, each containing a single repetition of a unique sound sequence.

Experimental manipulations

We manipulated several variables across each repetition of a unique sound sequence. At encoding, we used two sound sources: SG and EG, with half of the encoding sounds overall in the experiment belonging to each type. We counterbalanced the order of the two events within each sequence, thus half of the sequences began with EG sounds and then SG sounds and vice versa. Additionally, we manipulated the position of the test pairs at encoding. In half of the trials the test sounds appeared in the third and sixth positions of the encoding sequence, and the other half had them in the fourth and seventh positions. We also varied the length of the events, which ranged from 2 to 7 sounds. The varied lengths allowed us to counterbalance the appearance at encoding of both retrieval test pair conditions (within and across) in each test position. However, we included 24 catch trials (one for each block) where we asked about the fifth and eighth positions, to ensure that participants didn’t notice that we were asking about the same positions repeatedly. The retrieval phase of these catch trials was not included in the analysis.

At retrieval, we created two test pair conditions: within and across. In the within condition, half of the trials featured a SG-SG pair, while the other half featured an EG-EG pair. In the across condition, half of the trials presented a SG sound followed by an EG sound, and the other half presented an EG sound followed by a SG sound. Finally, we presented the test sounds in either the correct sequential order or the inverse order, with each version used in half of the trials.

Motor-only blocks

In self-generation studies, to be able to assess the effects of actions on auditory ERPs, it is standard procedure to eliminate the motor-related activity from the SG sound ERPs prior to comparing the auditory responses between SG and EG sounds (SanMiguel et al., 2013). To subtract the motor component from the SG sound responses in the ERP analysis, we added motor-only (M) blocks. These blocks were equal to the first 4 blocks of experimental trials but without a retrieval phase and without auditory stimuli. This ensured that the M evoked potentials were elicited by similar conditions to the SG sounds. These sequences were presented in 4 separate blocks of 9 sequences each, creating a total of 36 motor sequences which gave us a mean of 162 M items per participant (2 to 7 M items per sequence). Before the beginning of each motor block, we warned

participants that they were in a motor only block and that, consequently, no question or sounds would appear.

Apparatus

Binaural presentation of auditory stimuli was achieved using over-ear headphones (Sennheiser, HD 558). Participants' button presses and responses were recorded using a silent response pad (Korg nanoPAD2) to avoid interference with the auditory stimuli. The instructional visual stimulation was delivered through an AT Radeon HD 2400 monitor. The experimental setup and control were conducted using MATLAB R2017a (The Mathworks Inc., 2017) in conjunction with the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007) and the Eyelink add-in toolbox for eye tracker control (SR Research).

To record the EEG activity, we used a Neuroscan SynAmps RT amplifier and CURRY 8 software (NeuroScan, Compumedics). We recorded data from 64 channels placed according to the international extended 10–20 electrode system (Chatrian et al., 1985) by means of 64 Ag/AgCl electrodes inserted in a nylon cap (Quick-Cap, Compumedics). The reference electrode was placed at the nose, and we placed 4 additional electrodes above and below the left eye and in the outer canthi of the eyes to record the electrooculogram. We continually recorded the EEG activity at a sampling rate of 500 Hz. Impedance levels were maintained below 5 k Ω to ensure proper signal quality.

To obtain the pupillometry, we recorded the pupil's area from the left eye along horizontal and vertical gaze position at 1000 Hz sampling rate using an EyeLink 1000 desktop mount (SR Research). Pupil area was assessed using a center-of-mass algorithm in the centroid mode of the eye tracker.

Procedure

We conducted the experiment in an electrically and acoustically shielded room. Participants sat with their head placed on a chinrest approximately 60cm from the screen. To ensure that participants were adequately prepared for the task, they were given a practice block consisting of five trials of a modified version of the task where the sounds were pure tones, the encoding phase had only six sounds and test sounds were separated only by one sound. They were allowed to repeat this block as many times as needed to ensure they understood the task. The main experiment consisted of 24 blocks of experimental trials and four blocks of motor-only trials, with the first motor block presented before the first experimental block, and every six blocks thereafter. To prevent participant fatigue, short breaks were included at least every seven blocks. The experiment lasted approximately 90 minutes without preparation or breaks.

Behavioral analysis

To evaluate participants' order memory performance, we calculated the percentage of correctly answered retrieval questions ("is this the correct order?") for pairs of sounds that were presented either within or across events, and for each specific combination of sound source (EG-EG, SG-SG, SG-EG, EG-SG) separately. We conducted two-tailed t-tests to compare the percentage of correct responses between conditions and sound source combinations.

EEG preprocessing

We measured electrophysiological activity to retrieve event-related potentials (ERPs) in response to the encoding sounds and button presses in the motor-only (M) trials. The raw EEG data was bandpass filtered from 1 to 25 Hz using a Kaiser window (Kaiser β 5.653, filter order 1812) and manually examined to remove continuous atypical artifacts and to identify faulty electrodes. We then applied Independent Component Analysis (ICA) using the logistic infomax ICA algorithm (Onton & Makeig, 2006) to identify the eye-movement related components. The ICA weights of those components were subtracted from the raw bandpass filtered data. Remaining artifacts were rejected by applying a threshold of 75 μ V maximal signal-change per epoch and malfunctioning electrodes were interpolated spherically. Participants with excessive artifacts in

their EEG recordings ($> 60\%$ of epochs meeting the rejection criteria) were excluded. All the catch trials and the encoding trials with incorrect button presses were removed from the analysis. Epochs of -200 to +600 ms relative to sound or button press onsets, with a baseline correction from -200 to 0 ms, were extracted for each encoding item (EG, SG, and M) and for encoding sounds that occurred in the last position before a boundary (bB) and the boundary sounds (i.e. after a switch in the sound source, the first sound elicited by the new source, B), separately for EG (EGbB, EGB) and SG sounds (SGbB, SGB). All SG epochs were corrected for motor activity (SG-M) by subtracting the M epochs' averages to remove the activation resulting from motor actions. For simplicity, however, we will from now on refer to SG-M as SG.

ERP analysis

In order to investigate the self-generation effects, we compared the responses to EG sounds and the SG sounds by extracting the mean amplitudes of the N1 component (window: 80-120 ms) at Cz and joined mastoids ($N1_{\text{mast}}$), the P2 component (window: 150-250 ms) at Cz, and the N1 subcomponents Na (window: 80-100 ms) and Tb (window: 110-150 ms) at the joined temporal electrodes (T8 and T7), following the methodology described in SanMiguel et al. (2013). We used two-tailed t-tests to compare the amplitudes of EG and SG on all the defined components: N1, $N1_{\text{mast}}$, P2, Na and Tb.

To examine the potential boundary effect, we evaluated the sensory response to boundary sounds by extracting the mean amplitudes of the N1 and P2 components at Cz. To assess context change we planned to extract the mean amplitude in the P3 window at its maximum midline electrode (Pz). However, contrary to our expectations, we observed a negative instead of positive response at this spatiotemporal location. Moving forward, we will refer to this component as the parietal negative response (PNR, window: 392-432). We used 2x2 repeated measures ANOVAs with the factors Source (EG vs. SG sounds) and Position (B vs. bB sounds) to compare the amplitudes on all the defined components: N1, P2, and PNR.

We conducted a data-driven analysis to explore the possibility of detecting further effects that may not have been captured by the targeted ERP analyses (Maris & Oostenveld, 2007). To control for multiple comparisons and identify clusters with significant group differences in multiple dimensions (electrode-time samples), we used a non-parametric cluster-based permutation analysis. We defined neighboring sensors using a Delaunay triangulation over a 2D projection of the electrode montage and set a minimum of 2 nearby electrodes per cluster. This analysis was performed for all comparisons, and for each comparison, a two-tailed t-test was performed on the extracted values of each electrode-time sample of the epochs. The resulting amplitude values were assessed using the non-parametric Monte Carlo method with 10000 random partitions. Clusters were created by grouping adjacent electrode-time samples exceeding a significance level set to 0.05. The sum of individual t-statistics was calculated within each cluster to obtain a cluster-level statistic, and the significance level (p-value) was calculated using the non-parametric Monte Carlo method.

We used EEGLAB (Delorme & Makeig, 2004) for EEG preprocessing and ERP analysis, Eeprobe (ANT Neuro) for ERP visualization and mean amplitude extraction, and Fieldtrip (Oostenveld et al., 2011) for the data-driven analysis.

Pupillometry preprocessing

We measured the pupil diameter in response to all the encoding items following a similar pipeline as described in (Paraskevoudi & SanMiguel, 2023a), using Fieldtrip (Oostenveld et al., 2011). The EyeLink software detected missing data and blinks which were linearly interpolated using a time window from -100 to 100 ms. Missed blinks were identified using peak detection on the velocity of the pupil signal (Urai et al., 2017). If they occurred less than 250 ms apart they were joined as a single blink and the resulting blinks were then linearly interpolated. To reduce measurement noise not likely to originate from physiological sources, the interpolated pupil time series were filtered using a 0.05-4 Hz third-order Butterworth filter. The effect of blinks and saccades on the pupil response was estimated through a deconvolution analysis and removed from the data using linear regression (Knapen et al., 2016; Urai et al., 2017). We used the residual bandpass filtered pupil time series (z-scored) for the evoked analyses (Slooten et al., 2019). The data was

downsampled to 100 Hz and epoched from -0.5 to 1.5 ms with a baseline correction of 500 ms before the sound onset. The average evoked response for each participant was obtained for all the EG and SG sounds at encoding, and for the boundary and the before boundary sounds separated by sound source (EGB, EGbB, SGB, SGbB). To further explain the boundary pupil data we decided a posteriori to conduct an exploratory (not pre-registered) analysis to assess the overall pupil response for each event type, depending on the event order within a sequence (1st or 2nd). For this additional analysis, we aggregated all preprocessed (except for baseline correction) sound epochs belonging to the same event, and calculated the overall mean pupil diameter for the event. We then calculated the grand average for all first events (sounds before the boundary) and all second events (sounds after the boundary), separately for EG events and SG events. It is important to clarify that we did not perform any motor correction in the pupillometry analysis because our goal was to evaluate the impact of motor actions on the pupil response.

Pupillometry analysis

Non-parametric permutation statistics were used to test for group-level significance of the individual averages, following the same procedure as for the EEG data except that for pupillometry the clusters were defined based on the time dimension only. We tested for possible differences in evoked pupil responses computing t -values of the difference between the conditions of interest and thresholded these t values at a p value of .05. We tested for the main effects of Source (EG vs. SG sounds) and Position (B vs. bB sounds) as well as their interaction (performed on the difference waves EGB-EGbB vs SGB-SGbB). Adjacent time-samples that passed the threshold of the p -value ($< .05$) were constituted as clusters. The sum of individual t -statistics was calculated within each cluster to obtain a cluster-level statistic, and the significance level (p -value) was calculated using the non-parametric Monte Carlo method with 10000 random partitions. Additionally, to evaluate the effects in the overall event pupil response depending on event order we performed a 2x2 repeated measures ANOVA on Source (EG vs. SG) and Order (1st vs 2nd) of the two events that formed each sequence.

Correlations

To investigate our hypotheses on the relationship between the behavioral, electrophysiological, and pupillometry results we tested the possible correlations between these effects, considering only the physiological effects that were significant in the previous analyses. For the ERP responses and pupil data we included the self-generation effect (amplitude difference of EG minus SG) and the boundary effect separately by source (EGB minus EGbB; SGB minus SGbB) for each significant component or cluster. To obtain one value to perform the correlations on the significant clusters, we extracted the mean amplitude of the significant time window at its maximal electrode for the ERPs, and the peak of the difference wave for the pupil data. We then correlated the electrophysiological and neuromodulatory (pupil dilation) effects with the hypothetical production effects (the difference in temporal order memory performance between across and within), and directly with the memory performance in each combination of sound sources (EG-EG, SG-SG, SG-EG, EG-SG) using Pearson's correlation coefficients.

Results

1.

Behavioral

The present study examined the accuracy of participants' order memory by analyzing the mean percentage of correct responses to the retrieval question ("is this the correct order?") for sound pairs encoded within or across events, and for each sound source pair combination (EG-EG, SG-SG, SG-EG, EG-SG, Figure 2). Two-tailed t -tests revealed no significant difference in performance between sound pairs encoded within versus across events, $t(24) = -.966$, $p = .344$. Examining the within condition, there was no significant difference in performance for sound pairs comprised of two encoded EG sounds versus two encoded SG sounds, $t(24)$

$= .536$, $p = .597$. Additionally, for the across pairs, there was no significant difference in performance depending on whether the first sound was EG or SG, $t(24) = .324$, $p = .748$. These findings indicate that the participants' order memory performance was not influenced by the sound source at encoding nor by the presence of two encoding events.

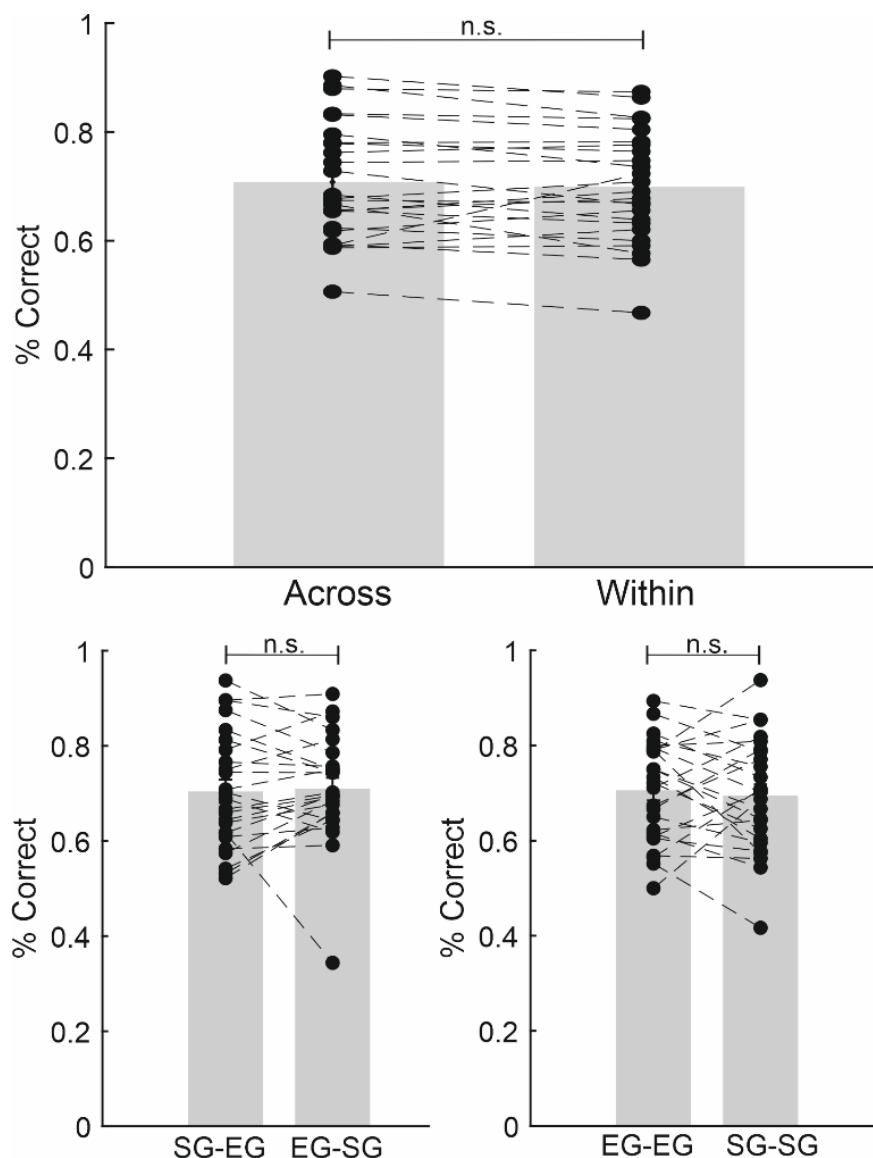


Figure 2 . Behavioral performance on temporal order memory. Bar plots illustrate both the within- and across-condition, while considering each pair combination of sound sources, externally-generated (EG) and self-generated (SG), during the encoding phase. Specifically, the across-condition is represented by the combinations SG-EG and EG-SG, while the within-condition is represented by EG-EG and SG-SG. Individual data points are depicted alongside the bar plots and connected by discontinuous lines.

Electrophysiological

Results of the targeted ERP analysis on the self-generation effects showed significant differences between EG and SG on several components (Figure 3). Specifically, the amplitude of the N1 component at Cz was

significantly suppressed for SG compared to EG ($t(24) = -6.671, p < .001, d = .101$). Conversely, $N1_{\text{mast}}$ did not differ significantly between EG and SG ($t(24) = .791, p = .437$). The Na and Tb components were also significantly modulated, with SG enhancement for Na and attenuation for Tb ($t(24) = 2.556, p = .017, d = .310$ and $t(24) = -2.597, p = .016, d = .397$, respectively). Moreover, the P2 amplitude at Cz was significantly suppressed for SG compared to EG ($t(24) = 6.985, p < .001, d = .984$). The cluster-based analysis showed more negative values over frontocentral electrodes for the EG compared to SG from 0 to 114 ms ($t = -2.453, p = .025$) and more positive values over frontocentral electrodes from 158 to 254 ms ($t = 3.784, p = .001$) and parietal electrodes from 528 to 598 ms ($t = 3.181, p = .007$), for the EG compared to SG. Thus, the negative cluster (0 to 114 ms) encompassed the Na (80-100 ms) and N1 (80-120 ms) components, and the first positive cluster (158 to 254 ms) encompassed the P2 (150-250 ms) component.

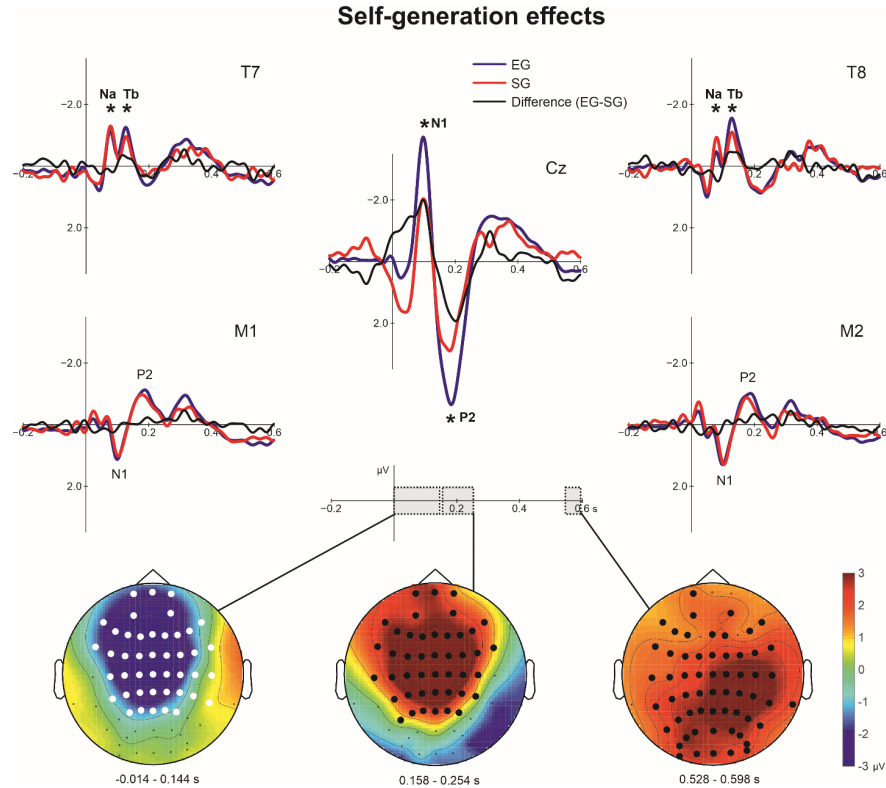


Figure 3. Electrophysiological results comparing the externally-generated (EG) and self-generated (SG) stimuli at encoding. The top section shows ERPs recorded on the analyzed electrodes. At Cz, M1, and M2, the analyzed components are N1 and P2, while at T7 and T8, the N1 subcomponents Na and Tb are examined. Significance is indicated by asterisks. The bottom section displays topographical plots representing three significant clusters identified through the ERP data-driven analysis comparing the EG vs the SG stimuli. The gray shading indicates the time windows corresponding to each cluster, the highlighted cluster electrodes indicate whether a cluster is negative (white) or positive (black) dots.

To examine the boundary effect, we examined responses to the boundary (B) and before-boundary (bB) sounds and performed 2x2 repeated measures ANOVAs (Source x Position) on N1 and P2 (at Cz), as well as on PNR at Pz to test for possible interactions between self-generation and boundary effects (Figure 4). Related to the N1, we obtained a main effect of Source ($F(1, 24) = 38.184, p < .001, \eta_p^2 = .614$) but no main effect of Position nor an interaction, thus reflecting the self-generation effect formerly reported, that is, a reduction of the N1 amplitude for SG sounds regardless of position. For the P2, we obtained a main effect of Source ($F(1, 24) = 20.201, p < .001, \eta_p^2 = .457$), as expected, and Position ($F(1, 24) = 22.657, p > .001$,

??2 = .486), which we did not predict, but no interaction. This indicates that P2 was significantly enhanced for boundary sounds maintaining the P2 attenuation for the SG sounds regardless of position. Finally, for the PNR there was a significant interaction between Source and Position ($F(1, 24) = 8.417, p = .008, \eta_p^2 = .260$). The post-hoc comparisons showed a significant effect of Position only for the SG sounds ($t(24) = -2.274, p = .032, d = 1.276$), indicating that the PNR was enhanced for the boundary SG sounds. However, there was no effect of Position for the EG sounds ($t(24) = .994, p = .330$). Thus, in terms of modulation of the PNR, starting to perform actions seemed to mark a clear boundary while stopping to perform actions did not. The cluster-based analysis revealed a significant interaction temporally overlapping the N1-P2 complex (window: .084 to .198 s, $t = 4643.049, p = .009$). Post-hoc comparisons revealed that SG sounds differed significantly depending on Position, where the boundary SG sounds elicited more positive responses compared to the before-boundary SG sounds (window: .084 to .198 s, $t = 6237.125, p < .001$); contrary, EG sound responses were not modulated by Position. The significant results from the cluster-based analysis overlapped with the N1 and P2 windows from the targeted ERP analysis, however, the cluster showed a posterior scalp distribution.

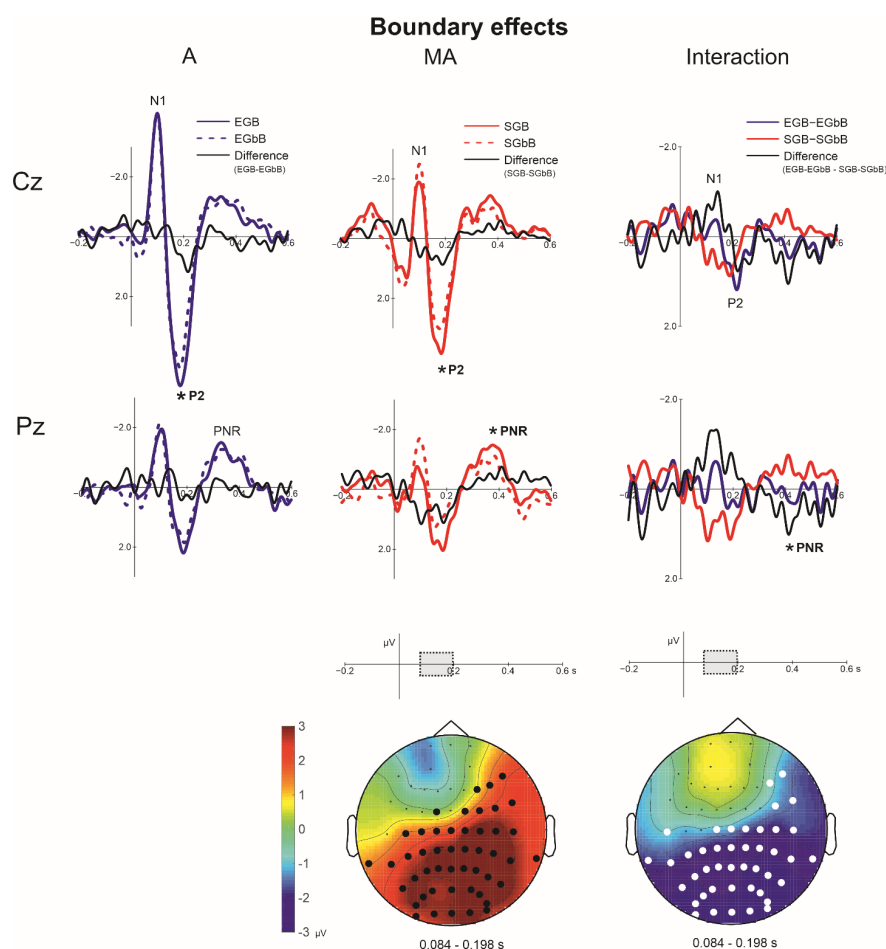


Figure 4. Electrophysiological results comparing the boundary effect (boundary, B, minus before-boundary, bB) on externally-generated (EG) and self-generated (SG) stimuli at encoding and the interaction between boundary and self-generation effects. The top section displays the recorded ERPs on the analyzed electrodes. At Cz, the N1 and P2 components were analyzed, while at Pz, the PNR was examined. The left-most and middle columns depict the boundary effect for the EG and SG sounds, respectively. In the rightmost column the difference waves depicting the boundary effects are compared between the SG and EG sounds (inter-

action). Asterisks indicate significant effects. The bottom section exhibits topographical plots representing the two significant clusters identified through the ERP data-driven analysis. The gray shading indicates the time windows corresponding to each cluster, and the electrodes included in each cluster are highlighted with dots, the color indicating whether the cluster is negative (white) or positive (black).

Pupillometry

The pupil analysis amongst all the encoding sounds showed a significant difference in pupil diameter between EG and SG sounds (Figure 5). Specifically, cluster-based permutation t-tests revealed a larger pupil diameter for SG sounds compared to EG sounds in the time window spanning from $-.130$ to $.890$ s ($t = -750.562, p < .001$).

Regarding the boundary effect, the pupil response cluster-based permutation t-tests revealed that the interaction between Position and Source was significant, (window: $-.230$ to 1.500 s, $t = -1170.038, p < .001$). Further analysis on the significant time window showed a simple main effect of Position on EG sounds (window: $.790$ to 1.100 s, $t = -590.941, p < .001$). This indicated that the phasic pupil response to EGB sounds was smaller than the response to EGbB sounds. Similarly, we found a simple main effect of Position on SG sounds (window: $-.240$ to 1.100 s, $t = 990.362, p < .001$). However, this indicated the opposite effect; the phasic pupil response to the SGB sounds was larger than to the SGbB sounds. Regarding Source, there was a simple main effect on before-boundary sounds (window: $.010$ to $.620$ s, $t = -310.524, p < .001$). This goes in accordance with the previous analysis amongst all the encoding sounds, showing enhanced pupil diameter for SGbB compared to EGbB sounds. Finally, there was a simple main effect of Source on boundary sounds (window: $-.220$ to 1.100 s, $t = -1131.452, p < .001$), indicating the same pattern of effect on the Position on SG sounds, that is, larger phasic response for the SGB compared to the EGB sounds. Overall, these findings suggest that both position and source play a significant role in the boundary effect on the pupil response. Additionally, the cluster-based analysis identified a difference on the baseline period (window: -0.490 to -0.260 s, $t = 169.787, p = .021$). However, due to the limited duration of the baseline period, we deemed it insufficient for drawing meaningful conclusions. Instead, considering that boundary effects on the baseline period of the boundary sound might relate to the direction of the boundary (from EG to SG or from SG to EG), to further explore this, we examined the differences in the overall pupil response for each event type taking event order into account, by conducting a 2x2 repeated measures ANOVA on Source (EG vs. SG) and Order (1st vs. 2nd event) of the two events comprising each sequence. The analysis revealed a significant interaction between Source and Order ($F(1, 24) = 10.619, p = .003, \eta_p^2 = .307$), main effect of Source ($F(1, 24) = 50.755, p < .001, \eta_p^2 = .679$) and Order ($F(1, 24) = 39.074, p < .001, \eta_p^2 = .619$). In terms of the simple main effects of Source, the overall event response was found to be larger for the SG events compared to the EG events during both the 1st event ($t(24) = -4.674, p < .001, d = .363$) and the 2nd event ($t(24) = -8.915, p < .001, d = .241$) of the sequence. Regarding the simple main effects of Order, 2nd events had significantly lower pupil dilations than 1st events both for EG events ($t(24) = 8.691, p < .001, d = .145$) and for SG events ($t(24) = -4.553, p = .048, d = .240$).

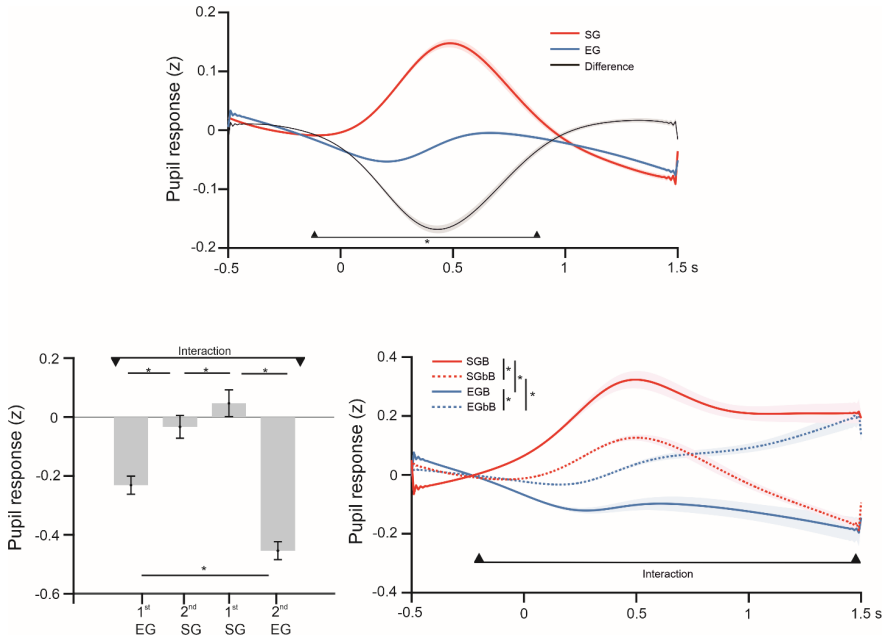


Figure 5. Pupil responses during encoding. The top panel shows the evoked pupil responses at encoding to all externally-generated (EG) and self-generated (SG) stimuli. The bottom right panel shows the evoked pupil responses for EG and SG depending on whether they were the last stimulus before a boundary (bB) or the boundary stimulus (B). The bottom left panel displays the overall mean pupil response of the entire events, separately for EG and SG events depending on order of presentation within a sequence (1st event or 2nd event). Error bars show the standard error of the mean (SEM). Significance is marked by a black line encompassing the significant time window.

Correlations

Table 1 displays the correlations conducted in this study to explore the possible relationships between the electrophysiological (ERPs) and neuromodulatory (pupil dilation) effects of actions at encoding and the memory performance effects at retrieval on the one hand; and the possible relationships between the two physiological effects of actions at encoding. The results indicate that the ERP SG effects (EG-SG) on each significant component did not exhibit correlations with the effect of actions on pupil dilation (EG-SG) or the effect of source on the performance of within pairs (EG_EG - SG_SG). Additionally, the pupil dilation effects did not demonstrate correlations with memory performance. Regarding the boundary effect, we examined the correlation between each significant physiological effect of the boundary (B-bB) and the corresponding associated memory performance for that boundary type. In cases where an interaction was present, we explored separately the boundary effect for each sound source (EG and SG). Notably, we discovered a significant negative correlation between the boundary effect on memory performance (within minus across) and the modulation of P2 for boundary sounds ($r = -.425$, $p = .034$). This finding suggests that larger performance differences between across and within events were associated with a smaller boundary effect at P2. However, no correlations were found between the behavioral boundary effect and the pupil, nor between the pupil and the ERP.

Correlations

		r	p
SG effects	SG effects		

Memory performance for within pairs (EG_EG – SG_SG)	N1 (EG - SG)	-.099	.638
	Na (EG - SG)	.133	.526
	Tb (EG - SG)	.035	.870
	P2 (EG - SG)	.204	.328
	Pupil (EG - SG)	.313	.127
Pupil (EG - SG)	N1 (EG - SG)	-.007	.972
	Na (EG - SG)	.070	.741
	Tb (EG - SG)	-.061	.773
	P2 (EG - SG)	.040	.848
Boundary effect	Boundary effect		
Memory performance (Within – Across)	P2 (B - bB)	-.425	.034
Memory performance (Within - Across EG-SG)	Parietal cluster (B - bB for SG)	-.260	.209
	PNR (B - bB for SG)	.030	.886
	Pupil (B - bB for SG)	-.209	.315
Memory performance (Within – Across SG-EG)	Pupil (B - bB for EG)	-.287	.164
Pupil (B - bB for EG)	P2 (B - bB)	-.218	.296
Pupil (B - bB for SG)	Parietal cluster (B - bB for SG)	-.256	.216
	P2 (B - bB)	.055	.792
	PNR (B - bB for SG)	-.236	.256

Table 1. Correlations between the memory performance, the significant self-generation and boundary effects on ERP components and the significant self-generation and boundary effects on pupil responses.

Discussion

The present study's primary aim was to explore the possible influence of motor actions to structure auditory memory. To do so, we manipulated the presence or absence of actions during the encoding of sound sequences and examined memory recall for the temporal order of presentation of sound pairs from the encoded sequences. We recorded electrophysiological and pupillary responses during the sound sequence encoding, aiming to relate the known physiological effects of actions during sensory processing to the possible effects of said actions on memory. We hypothesized that the neurophysiological processes engaged by actions, and the ensuing action-related modulation of sensory processing during encoding would promote the differentiated storage of self- and externally-generated sounds in memory. At the physiological level, we observed distinct modulations in both ERPs and pupil diameter in relation to the sounds encoded at the position coinciding with a change in sound source (SG or EG). This suggests that indeed the presence or absence of actions acted as a meaningful context for the sound encoding, and that specific neurophysiological mechanisms marked the processing of boundaries between the two different contexts. However, in contrast to our initial hypothesis, memory performance was not affected in any way by the actions performed during encoding. We conclude that the mere presence or absence of overt actions during sound sequence encoding, and the neurophysiological processes engaged by them, does not meaningfully structure auditory memory representations.

Additionally, this experiment contributes to our understanding of the role of motor actions in modulating the

strength of memory traces, more directly relating to the production effect literature. Previous studies testing the influence of actions on auditory memory encoding have reported either impaired, equal, or enhanced single-item recognition memory for sounds encoded concurrently to the execution of an action compared to sounds in isolation (Font-Alaminos et al., 2023; Paraskevoudi & SanMiguel, 2023a, 2023b). Importantly, in these studies, there was no contingent relationship between actions and sounds (Font-Alaminos et al., 2023; Paraskevoudi & SanMiguel, 2023a). Their findings suggest that the presence of action alone during sound encoding does not reliably contribute to the production effect, and thus that beyond action-sound coincidence, action-sound predictability may play a fundamental role in the memory advantage for self-generated sounds. In our study, we aimed to create conditions that more closely resemble those of studies reporting the production effect. This resulted in SG sounds being fully predictable in time, while EG sounds remained unpredictable, resembling natural conditions where external inputs are inherently unpredictable, but self-generated ones are not. This setup also resembles more closely the paradigms in which the production effect has been reported, where stimuli that receive a memory advantage are generated in the context of well-established action-effect relationships, such as using one’s own voice (MacLeod, 2011), or playing a musical instrument (Brown & Palmer, 2012). Nevertheless, in the present study the sound’s identity remained equally unpredictable for SG and EG sounds. Moreover, while the previous studies (Font-Alaminos et al., 2023; Paraskevoudi & SanMiguel, 2023a, 2023b) tested for effects of actions on single-item recognition memory, here, we explored possible effects on sequence memory, testing memory recall for sequential order. Despite these changes, we still could not demonstrate any effects of the presence of actions during sound encoding on memory performance. These results, summed to the previous findings, start to corroborate that actions alone, and the neurophysiological modulatory mechanisms associated with them, do not underlie or even seem to meaningfully contribute to the production effect, suggesting that additional factors beyond the mere presence of motor actions are critically involved in memory enhancement for SG stimuli. Considering the lack of feeling of control reported by several participants in the present study, the feeling of agency and control over the stimuli may be a key factor contributing to memory improvement, rather than the mere generation of a sound by a motor action during encoding. Recent studies have demonstrated that the presence of agency enhances the ability to remember the temporal order of events (Houser et al., 2022) and that voluntarily initiating the onset of stimuli improves working memory and speeds up visual and attentional processes (Loyola-Navarro et al., 2022). These findings suggest that the cognitive aspect of control and agency may play a crucial role in memory enhancement for self-generated stimuli (Sturm et al., 2023).

To delve into the effects of self-generation during encoding, we examined sensory ERP components. Consistent with well-established findings on predictable SG sounds, our study replicated the attenuated sensory responses to SG sounds on N1, Tb and P2 (Fu et al., 2006; Houde et al., 2002; Paraskevoudi & SanMiguel, 2023b). These results reinforce the notion that self-generated stimuli are processed differently at the sensory level compared to externally-generated stimuli. Notably, the absence of N1 attenuation at the mastoids suggests that, if any, the modulation of sound responses in areas located on the superior temporal plane (i.e., primary auditory cortex, which should be reflected at the mastoids, Horváth et al., 2012) is rather weak. Thus, the consistent lack of modulation at the mastoids (Font-Alaminos et al., 2023; Paraskevoudi & SanMiguel, 2023a, 2023b) supports the idea of concurrent modulation of sensory-specific and -unspecific components of the auditory N1 during the SG effect (Horváth et al., 2012; SanMiguel et al., 2013). Furthermore, we investigated the neuromodulatory effects surrounding motor actions by measuring pupil diameter during sound encoding. Our results revealed greater pupil dilation for SG stimuli compared to EG stimuli, indicating differential activity of the LC-NE system. This finding is in line with recent research highlighting the distinct modulation of pupil responses during the execution of goal-directed motor actions (Lubinus et al., 2022; Paraskevoudi & SanMiguel, 2023a, 2023b; Yebra et al., 2019). All in all, the physiological data shows the clear engagement of distinct neurophysiological processes during the processing of SG sounds, that could have substantial impact on perceptual and memory processes.

However, we did not observe any significant correlations between the electrophysiological, neuromodulatory effects and any memory performance differences between EG and SG sounds. This suggests that the observed modulations in sensory processing, as reflected by the attenuated sensory ERP components and the pupil

dilation for SG sounds, are not meaningfully modulating the strength of memory traces. Regarding the relationship between the different physiological effects observed during encoding, it is worth noting that previous research has demonstrated that larger pupil diameter for highly predictable self-produced sounds significantly correlates with greater suppression of the Tb component (Paraskevoudi & SanMiguel, 2023b). We could not replicate this correlation between the pupil and EEG measures. Although the two studies were similar in terms of temporal predictability conditions of SG and EG sounds, notably a significant correlation between Tb suppression and pupil dilation was found only when participants could have an additional sense of control over the stimulation, as they could choose out of several sound categories which one they wanted to produce on each instance (Paraskevoudi & SanMiguel, 2023b). All in all, sufficient converging negative evidence (Font-Alaminos et al., 2023; Paraskevoudi & SanMiguel, 2023a; and the present study) is starting to accumulate indicating that 1) actions alone, and the associated sensory attenuation reflected on ERPs and engagement of neuromodulatory mechanisms are not the basis for the production effect; and 2) increase in LC-NE activity engaged by making overt actions does not meaningfully contribute to attenuation of SG sound ERPs.

Still, actions could potentially meaningfully contribute to structure auditory memory. Returning to the main aim of this study, in contrast to our initial hypothesis, we did not find any significant differences in temporal order memory performance between the across and within conditions. In principle, this finding indicates that actions did not structure the encoding sequence in two differentiated events. However, recent research by Pu et al. (2022) tested principles of event segmentation manipulating several variables such as the length of events and position of the test items on a task where boundaries were defined as the change of color frame on a list of grey scaled pictures. Based on one of their observations they proposed the concept of the local primacy effect, which suggests that memory improvements, either across or within events, are strongest at the beginning of an event and gradually decrease as event positions move away from the event boundary. In our experiment, due to the need to control for the distance between the two test sounds in the encoding sequence, pairs of test sounds pertaining to the within event condition were primarily positioned at later local event positions, while pairs pertaining to the across events condition were positioned at earlier positions. This discrepancy in event position distribution may potentially explain our lack of differences between conditions. Furthermore, our findings are consistent with the results reported by Racciah et al. (2022) for the middle positions of temporal order memory as a function of serial position during encoding. In their study, where boundaries were defined as a change in male/female speaker over a list of spoken words, they also found no significant difference between across and within conditions on middle sequence positions. However, when they modeled for serial position to account for primacy and recency effects, they observed the expected segmentation effects.

Although we did not find differences in temporal order memory that could indicate event segmentation based on actions, our analysis of the electrophysiological data did reveal a boundary effect on the P2 component for both SG and EG boundary sounds, and additional boundary effects were observed only for SG boundary sounds (at EG-SG boundaries), specifically, an increased positivity over parietal cortex temporally overlapping the N1-P2 complex, followed by an increased late parietal negative response (PNR). This suggests that the processing of SG sounds is more sensitive to modulations on sensory processing related to changes in source-context within an encoding sequence. In other words, introducing actions, and engaging the associated neurophysiological effects, potentially acts as a stronger contextual shift than removing them. Interestingly, our analysis of pupil diameter showed an opposing pattern of boundary effects between sources (i.e., depending on the direction of the boundary), with SG boundary sounds (at EG-SG boundaries) eliciting the highest pupil diameter, EG boundary sounds (at SG-EG boundaries) eliciting the lowest, and both SG and EG sounds at the position immediately before the boundary falling in the middle range. Upon investigating the possible reasons for these contrasting effects, we observed EG boundary sounds were associated with larger pupil diameters at baseline compared to SG boundary sounds. According to the law of initial values (Lacey, 1956), the magnitude of a physiological response to a stimulus is influenced by the baseline level of the response. In the case of EG boundary sounds, it is possible that the pupil did not dilate significantly because, coming from a sequence of SG sounds, it was already relatively dilated, reaching a physiological

limit that hindered further dilation. Hence, when the baseline pupil size is larger, an equivalent dilatory response may have a reduced effect on the diameter, causing the dilation to appear smaller in comparison (Gilzenrat et al., 2010). To further understand these directional effects, we investigated the overall pupil response of the two events conforming a sequence. Although our paradigm and the exploratory analysis we performed does not exactly allow us to estimate the tonic response, previous literature on tonic and phasic pupil activity clearly shows an inverted U-shaped pattern, where the optimal phasic response is obtained at intermediate levels of tonic activity (McGinley et al., 2015; Yerkes & Dodson, 1908). Here, the boundary clearly occurs in contexts of different activity levels for one direction and the other, indicating an interaction between boundary and source which suggests that initiating actions marks more clearly the boundary than ceasing them which could explain why we didn't observe differences in temporal order memory.

Finally, examining the relationship between electrophysiological, neuromodulatory, and behavioral responses in relation to self-generation on boundary effects, we observed that only the ERP P2 component's boundary effect exhibited a correlation with the performance difference between within and across conditions. Nevertheless, we did not find any associations between pupil response and EEG indicating that although both demonstrated distinct boundary-related effects, they may be reflecting different underlying processes in terms of providing sequential structure. In conclusion, these findings imply that a change in context was indeed processed at the sensory level by the participants, however, it did not result in significant performance differences on temporal order memory.

Our study had several limitations that may explain the lack of significant behavioral results observed in the context of event segmentation paradigms on the visual modality and shed light on the complexity of event segmentation paradigms on auditory stimuli, particularly self-generated sounds. Firstly, the predictability of boundaries may have influenced participants' perception of events. We expected the motor act to be sufficient to signal a change in context and elicit an event boundary that could have consequences on memory. Event segmentation is an automatic process that occurs with little conscious control (Zacks & Swallow, 2007). However, it has been suggested that heightened attention is directed towards event boundaries due to their association with unpredictability and the optimization of information uptake (Kosie & Baldwin, 2019; Kurby & Zacks, 2008). Thus, the predictable nature of our paradigm's boundaries, as the sequence of actions to be performed was displayed at the beginning of the trial, may have hindered the necessary attentional engagement required to create distinct event boundaries.

Secondly, the length and repetition of events in our study could have impacted the results. Due to the constraints of designing an auditory temporal order memory task, we only utilized two repetitive event patterns, which limited the complexity and variability of the events. It remains unclear whether repeated pairing of two event types can lead to their integration into a more complex event type or if they are perceived as separate entities (Shin & DuBrow, 2021). Furthermore, while previous research demonstrated a robust (visual) boundary effect even with decreasing event length (Pu et al., 2022), we were not able to explore longer event lengths since pilot behavioral testing indicated a sharp drop on auditory memory performance overall for longer events. Therefore, including more varied event patterns could yield different results.

Another divergence from previous sequential order memory tasks was the absence of a specific task to assess associative memory for the source of the items. Unlike other studies, our focus was solely on examining the effects of the change in context in the form of a motor action. Thus, the inclusion of an associative memory task could have biased our results due to the increased attention on the contextual environment. Future research should explore this aspect to gain a more comprehensive understanding. Additionally, the dual task nature of our experiment, with participants simultaneously performing a motor task to generate sounds, may have acted as a distractor. Divided attention between the motor task and memory encoding can interfere with episodic memory processes, particularly memory for temporal order. Previous studies have demonstrated that memory for temporal order requires greater attentional resources and strategic processing than memory for individual items (Mangels et al., 2001; Troyer & Craik, 2000).

In conclusion, our study investigated the influence of motor actions on memory encoding by examining order memory performance and electrophysiological and pupillary responses during sequence encoding. While

the behavioral findings did not reveal significant differences in order memory performance based on event segmentation, we observed distinct modulations in sensory processing and pupil dilation related to motor actions. These findings contribute to our understanding of the production effect and suggest that factors beyond motor actions, such as the feeling of agency, may play a crucial role in memory enhancement for self-generated stimuli. Furthermore, the study’s limitations, highlight the complexity of event segmentation on auditory paradigms and the need for future research to deepen our understanding of the underlying mechanisms and the role of motor activity in shaping our experiences.

References

- Aston-Jones, G., & Cohen, J. D. (2005). An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience* , 28 (1), 403–450. <https://doi.org/10.1146/annurev.neuro.28.061604.135709>
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* , 403 (6767), 309–312. <https://doi.org/10.1038/35002078>
- Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience* , 1 (7), 635–640. <https://doi.org/10.1038/2870>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision* , 10 (4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Brown, R. M., & Palmer, C. (2012). Auditory–motor learning influences auditory memory for music. *Memory & Cognition* , 40 (4), 567–578. <https://doi.org/10.3758/s13421-011-0177-x>
- Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten Percent Electrode System for Topographic Studies of Spontaneous and Evoked EEG Activities. *American Journal of EEG Technology* , 25 (2), 83–92. <https://doi.org/10.1080/00029238.1985.11080163>
- Clewett, D., DuBrow, S., & Davachi, L. (2019). Transcending time in the brain: How event memories are constructed from experience. *Hippocampus* , 29 (3), 162–183. <https://doi.org/10.1002/hipo.23074>
- 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. <https://doi.org/10.1038/s41467-020-17851-9>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* , 134 (1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences* , 11 (3), 357–374. <https://doi.org/10.1017/S0140525X00058027>
- 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), 1277–1286. <https://doi.org/10.1037/a0034024>
- DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of Learning and Memory* , 134 , 107–114. <https://doi.org/10.1016/j.nlm.2016.07.011>
- Eggermann, E., Kremer, Y., Crochet, S., & Petersen, C. C. H. (2014). Cholinergic Signals in Mouse Barrel Cortex during Active Whisker Sensing. *Cell Reports* , 9 (5), 1654–1660. <https://doi.org/10.1016/j.celrep.2014.11.005>
- Font-Alaminos, M., Paraskevoudi, N., & SanMiguel, I. (2023). Actions do not clearly impact auditory memory. *Frontiers in Human Neuroscience* , 17 , 1124784. <https://doi.org/10.3389/fnhum.2023.1124784>

- Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society B: Biological Sciences* , 355 (1404), 1771–1788. <https://doi.org/10.1098/rstb.2000.0734>
- Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro, E., Jr, Andrew, C. M., Yaguez, L., van Haren, N. E. M., Matsumoto, K., & McGuire, P. K. (2006). An fMRI Study of Verbal Self-monitoring: Neural Correlates of Auditory Verbal Feedback. *Cerebral Cortex* , 16 (7), 969–977. <https://doi.org/10.1093/cercor/bhj039>
- Gagl, B., Hawelka, S., & Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: Analysis and correction. *Behavior Research Methods* , 43 (4), 1171–1181. <https://doi.org/10.3758/s13428-011-0109-5>
- Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience* , 10 (2), 252–269. <https://doi.org/10.3758/CABN.10.2.252>
- Gygi, B., & Shafiro, V. (2010). Development of the Database for Environmental Sound Research and Application (DESRA): Design, Functionality, and Retrieval Considerations. *EURASIP Journal on Audio, Speech, and Music Processing* , 2010 , 654914. <https://doi.org/10.1155/2010/654914>
- Hocking, J., Dzafic, I., Kazovsky, M., & Copland, D. A. (2013). NESSTI: Norms for Environmental Sound Stimuli. *PLOS ONE* , 8 (9), e73382. <https://doi.org/10.1371/journal.pone.0073382>
- Horvath, J., Maess, B., Baess, P., & Toth, A. (2012). Action-sound coincidences suppress evoked responses of the human auditory cortex in EEG and MEG. *Journal of Cognitive Neuroscience* , 24 (9), 1919–1931. <https://doi.org/10.1162/jocn.a.00215>
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the Auditory Cortex during Speech: An MEG Study. *Journal of Cognitive Neuroscience* , 14 (8), 1125–1138. <https://doi.org/10.1162/089892902760807140>
- Houser, T. M., Tompary, A., & Murty, V. P. (2022). Agency enhances temporal order memory in an interactive exploration game. *Psychonomic Bulletin & Review* , 29 (6), 2219–2228. <https://doi.org/10.3758/s13423-022-02152-7>
- Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin* , 139 (1), 133–151. <https://doi.org/10.1037/a0028566>
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What’s new in psychtoolbox-3. *Perception* ,36 (14), 1–16.
- Knapen, T., Gee, J. W. de, Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and Ocular Factors Jointly Determine Pupil Responses under Equiluminance. *PLOS ONE* , 11 (5), e0155574. <https://doi.org/10.1371/journal.pone.0155574>
- Kosie, J. E., & Baldwin, D. (2019). Attentional profiles linked to event segmentation are robust to missing information. *Cognitive Research: Principles and Implications* , 4 (1), 8. <https://doi.org/10.1186/s41235-019-0157-4>
- 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>
- Lacey, J. I. (1956). The Evaluation of Autonomic Responses: Toward a General Solution. *Annals of the New York Academy of Sciences* ,67 (5), 125–163. <https://doi.org/10.1111/j.1749-6632.1956.tb46040.x>
- Loyola-Navarro, R., Moenne-Loccoz, C., Vergara, R. C., Hyafil, A., Aboitiz, F., & Maldonado, P. E. (2022). Voluntary self-initiation of the stimuli onset improves working memory and accelerates visual and attentional

- p processing.
- Heliyon*
- , 8 (12), e12215.
- <https://doi.org/10.1016/j.heliyon.2022.e12215>
- Lubinus, C., Einhauser, W., Schiller, F., Kircher, T., Straube, B., & van Kemenade, B. M. (2022). Action-based predictions affect visual perception, neural processing, and pupil size, regardless of temporal predictability. *NeuroImage* , 263 , 119601. <https://doi.org/10.1016/j.neuroimage.2022.119601>
- MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica* , 98 (2–3), 291–310. [https://doi.org/10.1016/s0001-6918\(97\)00047-4](https://doi.org/10.1016/s0001-6918(97)00047-4)
- MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review* , 18 (6), 1197–1202. <https://doi.org/10.3758/s13423-011-0168-8>
- MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science* , 26 , 390–395. <https://doi.org/10.1177/0963721417691356>
- Mangels, J. A., Picton, T. W., & Craik, F. I. M. (2001). Attention and successful episodic encoding: An event-related potential study. *Cognitive Brain Research* , 11 (1), 77–95. [https://doi.org/10.1016/S0926-6410\(00\)00066-5](https://doi.org/10.1016/S0926-6410(00)00066-5)
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* , 164 (1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- McCurdy, M. P., Leach, R. C., & Leshikar, E. D. (2017). The generation effect revisited: Fewer generation constraints enhances item and context memory. *Journal of Memory and Language* , 92 , 202–216. <https://doi.org/10.1016/j.jml.2016.06.007>
- McGinley, M. J., David, S. V., & McCormick, D. A. (2015). Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron* , 87 (1), 179–192. <https://doi.org/10.1016/j.neuron.2015.05.038>
- Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology* , 12 , 436–450. [https://doi.org/10.1016/0022-1031\(76\)90076-7](https://doi.org/10.1016/0022-1031(76)90076-7)
- Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron* , 88 (6), 1281–1296. <https://doi.org/10.1016/j.neuron.2015.11.035>
- Onton, J., & Makeig, S. (2006). Information-based modeling of event-related brain dynamics. *Progress in Brain Research* , 159 , 99–120. [https://doi.org/10.1016/S0079-6123\(06\)59007-7](https://doi.org/10.1016/S0079-6123(06)59007-7)
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience* , 2011 . <https://doi.org/10.1155/2011/156869>
- Ozubko, J. D., & Macleod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology. Learning, Memory, and Cognition* , 36 (6), 1543–1547. <https://doi.org/10.1037/a0020604>
- Paraskevoudi, N., & SanMiguel, I. (2023a). Sensory suppression and increased neuromodulation during actions disrupt memory encoding of unpredictable self-initiated stimuli. *Psychophysiology* , 60 (1), e14156. <https://doi.org/10.1111/psyp.14156>
- Paraskevoudi, N., & SanMiguel, I. (2023b). *The multifaceted nature of self-generation effects: Independent and interactive effects of actions, predictability, and delay on sensory processing and memory encoding of sounds*. [Preprint]. Preprints. <https://doi.org/10.22541/au.168322155.53723594/v1>
- Pettijohn, K. A., Thompson, A. N., Tamplin, A. K., Krawietz, S. A., & Radvansky, G. A. (2016). Event boundaries and memory improvement. *Cognition* , 148 , 136–144. <https://doi.org/10.1016/j.cognition.2015.12.013>

- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology* , 118 (10), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>
- 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), 622. <https://doi.org/10.1038/s41467-022-28216-9>
- Raccah, O., Doelling, K. B., Davachi, L., & Poeppel, D. (2022). Acoustic features drive event segmentation in speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition* . <https://doi.org/10.1037/xlm0001150>
- Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences* , 17 , 133–140. <https://doi.org/10.1016/j.cobeha.2017.08.006>
- SanMiguel, I., Todd, J., & Schroger, E. (2013). Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. *Psychophysiology* ,50 (4), 334–343. <https://doi.org/10.1111/psyp.12024>
- Schroger, E., Marzecova, A., & Sanmiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *European Journal of Neuroscience* , 41 (5), 641–664. <https://doi.org/10.1111/ejn.12816>
- Shin, Y. S., & DuBrow, S. (2021). Structuring Memory Through Inference-Based Event Segmentation. *Topics in Cognitive Science* ,13 (1), 106–127. <https://doi.org/10.1111/tops.12505>
- Silva, M., Baldassano, C., & Fuentemilla, L. (2019). Rapid memory reactivation at movie event boundaries promotes episodic encoding. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* , July . <https://doi.org/10.1523/JNEUROSCI.0360-19.2019>
- Slooten, J. C. V., Jahfari, S., Knapen, T., & Theeuwes, J. (2019). Correction: How pupil responses track value-based decision-making during and after reinforcement learning. *PLOS Computational Biology* ,15 (5), e1007031. <https://doi.org/10.1371/journal.pcbi.1007031>
- Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cognitive, Affective, & Behavioral Neuroscience* , 3 (4), 335–345. <https://doi.org/10.3758/CABN.3.4.335>
- Sturm, S., Costa-Faidella, J., & SanMiguel, I. (2023). Neural signatures of memory gain through active exploration in an oculomotor-auditory learning task. *Psychophysiology* , e14337. <https://doi.org/10.1111/psyp.14337>
- Swallow, M. K., Jeffrey, Zacks. M., & Richard, Abrams. A. (2010). Event Boundaries in Perception Affect Memory Encoding and Updating. *J Exp Psychol Gen.* , 138 (2), p236. <https://doi.org/10.1037/a0015631>
- Troyer, A. K., & Craik, F. I. M. (2000). The effect of divided attention on memory for items and their context. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Experimentale* , 54 (3), 161–171. <https://doi.org/10.1037/h0087338>
- Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications* , 8 (1), 14637. <https://doi.org/10.1038/ncomms14637>
- Vinck, M., Batista-Brito, R., Knoblich, U., & Cardin, J. A. (2015). Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron* , 86 (3), 740–754. <https://doi.org/10.1016/j.neuron.2015.03.028>
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An Internal Model for Sensorimotor Integration. *Science* , 269 (5232), 1880–1882. <https://doi.org/10.1126/science.7569931>

Yebra, M., Galarza-Vallejo, A., Soto-Leon, V., Gonzalez-Rosa, J. J., de Berker, A. O., Bestmann, S., Oliviero, A., Kroes, M. C. W., & Strange, B. A. (2019). Action boosts episodic memory encoding in humans via engagement of a noradrenergic system. *Nature Communications* ,10 (1), 1–12. <https://doi.org/10.1038/s41467-019-11358-8>

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