The Costs (and Benefits?) of Effortful Listening for Older Adults: Insights from Simultaneous Electrophysiology, Pupillometry, and Memory

Jack W Silcox, Karen Bennett, Allyson Copeland, Sarah Hargus Ferguson, and Brennan R. Payne

Abstract

■ Although the impact of acoustic challenge on speech processing and memory increases as a person ages, older adults may engage in strategies that help them compensate for these demands. In the current preregistered study, older adults (n = 48) listened to sentences—presented in quiet or in noise—that were high constraint with either expected or unexpected endings or were low constraint with unexpected endings. Pupillometry and EEG were simultaneously recorded, and subsequent sentence recognition and word recall were measured. Like young adults in prior work, we found that noise led to increases in pupil size, delayed and reduced ERP responses, and decreased recall for unexpected words. However, in contrast to prior work in young adults where a larger pupillary response predicted a recovery of the N400 at the cost of poorer memory performance in noise, older adults did not show an associated recovery of the N400 despite decreased memory performance. Instead, we found that in quiet, increases in pupil size were associated with delays in N400 onset latencies and increased recognition memory performance. In conclusion, we found that transient variation in pupil-linked arousal predicted trade-offs between real-time lexical processing and memory that emerged at lower levels of task demand in aging. Moreover, with increased acoustic challenge, older adults still exhibited costs associated with transient increases in arousal without the corresponding benefits. ■

INTRODUCTION

Comprehending speech often feels effortless. However, speech perception and comprehension become increasingly effortful in acoustically challenging environments, for example, arising from competing speakers or background noise (Peelle, 2018). To overcome difficulties processing perceptually challenging speech, listeners allocate additional cognitive and neural resources to meet the increased demands, a phenomenon referred to as *listening effort* (Pichora-Fuller et al., 2016). Importantly, listening effort and its effects on speech processing appear to compound with age (Kuchinsky & Vaden, 2020; Degeest, Keppler, & Corthals, 2015). Yet our understanding of how acoustic challenge and effort impact higher-level comprehension processes (e.g., context processing) in the brain is currently underspecified among older adults. For example, several behavioral studies have shown that older adults may rely on supportive context to help overcome difficulties in speech perception relative to younger adults, whereas electrophysiological evidence suggests that older adults have a limited ability to take advantage of syntactic and semantic information from supportive contexts (for a review, see Payne & Silcox, 2019). Therefore, our understanding of how effort and acoustic challenge impact context use as a person ages is still incomplete. The goal of the current study was to adopt a multimodal approach (electrophysiology, pupillometry, and behavior) to elucidate how older adults use contextual information during real-time speech processing under acoustically challenging conditions.

Context Use and Aging

Speech comprehension requires the rapid and continuous processing of acoustic input, mapping of sounds onto lexical features, and integration of words into a message-level semantic representation. How do listeners accomplish this, given the incredible speed of natural speech (> 200 words per minute; Miller, Grosjean, & Lomanto, 1984)? One strategy listeners may use is to rely on top–down contextual constraints to form predictions about likely upcoming linguistic input (Federmeier, 2007, 2022; Brown & Kuperberg, 2015; Huettig, 2015). Indeed, over the last 70 years (Howes & Osgood, 1954; Taylor, 1953; Black, 1952; Miller, 1951), studies have explored the effects of prior linguistic context on perception and comprehension and have reliably shown that words congruent with (and more predictable from) prior context are

---

University of Utah

Journal of Cognitive Neuroscience 36:6, pp. 997–1020
https://doi.org/10.1162/jocn_a_02161

facilitated in processing relative to less expected input. Predictable words show faster visual and auditory word recognition in lexical decision tasks (Schwanenflugel & Shoben, 1985; Fischler & Bloom, 1979), shorter fixation durations during natural reading (Staub, 2015), improved word recognition in noise (Pichora-Fuller, Schneider, & Daneman, 1995), improved memory (Silcox, Mickey, & Payne, 2023; Silcox & Payne, 2021; Gordon-Salant & Fitzgibbons, 1997; Kutas, 1993), and facilitated semantic processing as revealed by recordings of ERPs (Federmeier, 2022; Kutas & Federmeier, 2011).

Despite robust effects of prior context on language processing, the literature on how context processing may change in older adulthood varies considerably (for a recent review, see Payne & Silcox, 2019). For example, speech perception studies—typically using offline measures (e.g., memory outcomes, word recognition performance)—find that older adults generally perform worse than younger adults when listening to speech in background noise. However, when that speech in noise uses linguistic context that supports prediction and integration, older adults tend to perform just as well as younger adults (Benichov, Cox, Tun, & Wingfield, 2012; Sheldon, Pichora-Fuller, & Schneider, 2008; Goy, Pichora-Fuller, van Lieshout, Singh, & Schneider, 2007; Pichora-Fuller, Schneider, MacDonald, Pass, & Brown, 2007; Murphy, Craik, Li, & Schneider, 2000; Gordon-Salant & Fitzgibbons, 1997). For example, Pichora-Fuller and colleagues (1995) examined word recognition accuracy as a function of contextual constraint and background noise in younger and older listeners. For less-constraining sentence contexts, older adults performed worse than younger adults on a word recognition task, especially as the background noise increased. When speech had highly constraining contexts that better supported prediction, both younger and older listeners performed better than they did when supportive context was absent. However, the increase in performance from low- to high-context conditions by older adults was much larger, such that any differences in performance between age groups on word recognition were eliminated. This suggests that not only is older adults' ability to successfully utilize linguistic context intact but older adults may stand to gain more from supportive, linguistic context than younger adults in part because of their worse performance in the absence of contextual constraints.

In contrast, electrophysiological research has consistently found age-related deficits in real-time context use during sentence comprehension, primarily reflected in differences in the N400 ERP component. The N400 is a centro-parietal negative deflection in the ERP waveform peaking at approximately 400 msec (with neural generators in superior and middle temporal gyrus, angular gyrus, anterior temporal cortex, and left inferior frontal cortex; Lau, Phillips, & Poeppel, 2008; Van Petten & Luka, 2006). Decades of work indicate that the N400 reflects early access to long-term semantic memory, with the magnitude of the N400 scaling with the amount of new semantic information activated in response to meaning-bearing stimuli (Federmeier, 2022; Brouwer, Crocker, Venhuizen, & Hoeks, 2017; Federmeier & Laszlo, 2009). Although younger adults show substantial facilitation in the N400 for words that are more predictable based on prior context, older adults consistently show reduced and delayed effects of context on the N400 (for reviews, see Payne & Silcox, 2019; Wlotko, Lee, & Fedeermeier, 2010), even in the absence of any perceptual challenges. Work by Kemmotsu and colleagues (2012) using magnetoencephalography and diffusion tensor imaging suggest that age-related N400 reductions are related to structural changes in the efficacy of long-range white matter pathways, including the uncinate fasciculus, a fiber tract implicated in semantic retrieval in language processing (Von Der Heide, Skipper, Klobusicky, & Olson, 2013). Critically, studies have shown that these age-related effects on the N400 appear to be selective to context processing, as lexical N400 effects such as lexical associative priming, word frequency, or orthographic neighborhood appear to be preserved or are even larger in aging (Jongman & Federmeier, 2022; Payne & Federmeier, 2018; Federmeier, Van Petten, Schwartz, & Kutas, 2003). Payne and Silcox (2019) reviewed the discrepant literatures on aging and context processing and identified substantial differences across studies in terms of task demands (e.g., presence and degree of perceptual load), modality (e.g., reading vs. listening), and outcome measures (e.g., real-time vs. offline; behavioral vs. neural) that mostly fell along paradigmatic traditions within specific fields (e.g., psycholinguistics, audiology, cognitive neuroscience). They argued that the key to developing a generalized understanding of the mechanisms of context use in aging required adopting a multimethod approach cutting across these largely siloed literatures.

Acoustic Challenge and Listening Effort

Listening effort is a complex and multidimensional construct (Alhanbali, Dawes, Millman, & Munro, 2019) reflecting more than mere acoustic demand (Winn & Teece, 2021). When obstacles to perception are present (e.g., trying to listen to a friend in a noisy café), more neural resources may be required to process speech at normal performance levels. Listening effort is thus more about how the listener responds to the demands of the listening task, rather than the demands of the task itself. In an experimental setting, the amount of effort experienced may vary from trial to trial and from participant to participant. This variability likely reflects both the acoustic challenge of the task and the arousal, fatigue, and motivation levels of a participant. Furthermore, the additional cognitive and neural resources needed to overcome increases in an acoustic challenge likely do not originate from a single "pool" of resources that all processes harness. Rogers and

Peelle (2022) reviewed the functional neuroimaging literature and showed that, depending on listener goals and priorities, listeners appear to strategically recruit a number of dissociable brain networks outside of the core perisylvian language network to help overcome acoustic demands (see also Peelle, 2018). Importantly, the engagement of these additional domain-general resources to assist in speech perception comes at the cost of other downstream processes. For example, it has been frequently observed that although increased listening effort can be deployed to maintain speech recognition performance with moderate increases in acoustic challenge, this consistently comes at the cost of poorer subsequent memory for that same speech (Payne et al., 2022; Silcox & Payne, 2021; Guang, Lefkowitz, Dillman-Hasso, Brown, & Strand, 2020; Piquado, Benichov, Brownell, & Wingfield, 2012; McCoy et al., 2005).

To understand how listening effort impacts processing, we need measures beyond average performance on a listening task. Francis and Love (2019) argued there is a "certain circularity" to using task performance "both as an independent factor related to effort (this task is more effortful because people perform worse on it), and as a dependent measure of the effort that participants devote to the task (people perform worse on this task because it is the more effortful one)" (p. 2) (see also van der Wel & van Steenbergen, 2018; Shenhav et al., 2017; Navon, 1984). To counter this circularity, researchers have been exploring alternative independent measures of effort. *Pupillometry* is one of several physiological measures that cognitive audiology has identified that may measure different aspects of listening effort (for a recent review, see Richter et al., 2023). Pupillometry is the continuous measurement of pupil size across time (for reviews, see Laeng & Alnaes, 2019; Mathôt, 2018; van der Wel & van Steenbergen, 2018; Sirois & Brisson, 2014). Most agree that there are two general categories of factors that determine pupil size: low level and higher level. Low-level factors include light level (via the pupil light/dark reflex) and focal distance (via the lens accommodation reflex), and when they are controlled for, pupil size is consistently sensitive to higher-level cognitive factors. For example, under constant lighting conditions, when working on difficult mental arithmetic problems (Sirois & Brisson, 2014; Hess & Polt, 1964), tasks that require cognitive control (for a review, see van der Wel & van Steenbergen, 2018), or listening tasks involving acoustic challenge (McGarrigle, Dawes, Stewart, Kuchinsky, & Munro, 2017; Miles et al., 2017; Wendt, Hietkamp, & Lunner, 2017; Koelewijn, de Kluiver, Shinn-Cunningham, Zekveld, & Kramer, 2015; Winn, Edwards, & Litovsky, 2015; Koelewijn, Shinn-Cunningham, Zekveld, & Kramer, 2014; Zekveld & Kramer, 2014; Kuchinsky et al., 2013; Koelewijn, Zekveld, Festen, & Kramer, 2012; Zekveld, Kramer, & Festen, 2010; for a review, see Zekveld, Koelewijn, & Kramer, 2018), the pupil typically increases with increasing difficulty. Although there is still an ongoing debate on the underlying neurophysiology driving these higher-level changes in pupil size, the locus coeruleus norepinephrine (LC-NE) arousal system, the superior colliculus orienting system, and the executive function network have all been proposed to modulate pupil size (for recent proposals, see Strauch, Wang, Einhäuser, Van der Stigchel, & Naber, 2022; Joshi & Gold, 2020).

Evidence suggests the task-evoked cognitive pupillary response is not merely an indication of task demand, but instead reflects some aspect of the effort experienced by a listener. For example, Wendt, Koelewijn, Ksiażek, Kramer, and Lunner (2018) had participants listen to and immediately recall sentences that were under continuously varying degrees of acoustic challenge. As the signal-to-noise ratio (SNR) worsened, speech recognition performance decreased and the recorded pupillary response increased. However, once performance dropped below 70% correct recognition, peak pupil size began to decrease with decreasing intelligibility. Wendt and colleagues (2018) found that the cognitive pupillary response had a nonlinear, inverted-U shape relationship with speech recognition performance (see also Ohlenforst et al., 2017, 2018; McMahon et al., 2016; Zekveld & Kramer, 2014). They argued that as speech initially became less unintelligible, the amount of effort required for successful perception increased, indicated by an increase in pupil size. However, as the speech input became more and more degraded, the likelihood of successful perception decreased so much that the listener's motivation to continue to engage in the task declined. As their motivation dropped, the amount of effort they were willing to allocate would likewise decrease. Therefore, the pupillary response represents a useful, independent, and continuous measure that indexes not how demanding a task is but rather how a listener responds to that demand.

Multimodal Assessment of Listening Effort and Context Processing

We know relatively little about how variation in effort and arousal impacts context processing and the N400 in speech comprehension. Silcox and Payne (2021) addressed this question directly by co-registering pupillometry and language-related ERPs while young normal-hearing adults listened to sentences in quiet and in noise. Participants listened to highly constraining sentence contexts completed with either expected or unexpected sentence-final words (e.g., "*The prisoners were planning their escape/party*") or low-constraint control sentences with unexpected sentence-final words (e.g., "*All day she thought about the party*.") Replicating prior work, Silcox and Payne (2021) showed increases in background noise led to increased pupil dilation (Zekveld et al., 2010), as well as a reduced context-based N400 effect on sentence-final target words (Obleser & Kotz, 2011). This suggests that acoustic challenge (on average) leads to increased effort and reduced context-based facilitation to semantic access. However, there was considerable trial-to-trial variability in both the N400 and pupil dilation to speech in noise. To examine how variation in pupil-linked arousal predicted context processing, the authors tested for single-trial coupling between the N400 and pupil dilation. For trials presented in noise, increased pupil dilation predicted larger N400 context effects. Thus, although noise impaired context processing on average, when listeners responded with increased arousal (reflected by increased pupil size), they were able to buffer against the negative effects of acoustic challenge, in essence "rescuing" the N400 context effect that was diminished by noise. However, this effort induced N400 recovery had downstream costs. For offline measures of word recall and sentence recognition, a larger pupil dilation response during speech listening was associated with poorer subsequent word recall and sentence recognition memory. Collectively, these findings suggest an effort-driven resource trade-off between real-time, context-based word processing and subsequent memory (see also Peelle, 2018; Zekveld et al., 2018; Pichora-Fuller et al., 2016; Rabbitt, 1968, 1991). As listeners exerted more effort to support online word recognition processes in the face of degraded speech (as observed by the increased N400 effect) fewer resources were available to allocate to memory encoding processes, leading to poorer long-term memory for speech. The multimodal measurement and analysis of pupil dilation, brain activity, and behavior made it possible to track these dynamic variations in effort and context use.

The Current Study

Silcox and Payne (2021) directly assessed acoustic challenge and arousal-based variation in context processing and memory, but only in young normal-hearing adults. To our knowledge, no prior work has examined the effects of background noise on context processing using both ERPs and memory outcomes. Moreover, although a growing literature has characterized the pupillary effort response to speech in noise in older adults (e.g., McGarrigle, Knight, Rakusen, Geller, & Mattys, 2021; Zhao, Bury, Milne, & Chait, 2019; Ayasse, Lash, & Wingfield, 2017), no work has examined the consequences of variation in effort and arousal for high-level language comprehension (such as context processing). Thus, the goal of the current study was to replicate the work of Silcox and Payne (2021) in a sample of older adults, a group for whom effort allocation and context processing are likely to vary substantially from young normal-hearing adults. Importantly, in addition to detailing the overall effects of acoustic challenge (e.g., background noise) on electrophysiological responses, pupil dilation, and memory, we additionally examine the trial-to-trial relationships between the task-evoked pupil dilation response (as a marker of trial-to-trial variation in the effort-related arousal response) and both electrophysiological responses and memory measures to better characterize how dynamic variation in effort and arousal affects online (i.e., real-time) context processing and subsequent memory.

METHODS

Preregistration and Data Availability Statement

This study was preregistered on the Open Science Framework website (https://osf.io/7p5r6). This preregistration provides justification of sample size and the a priori analysis plan. In general, this study is a replication of Silcox and Payne (2021) in older adults and follows the same general methodological and analytical plan seen in that study. Throughout the Methods and Results sections, we are explicit in which hypotheses and analyses were confirmatory in nature and which were exploratory (Nosek, Ebersole, DeHaven, & Mellor, 2018), and any and all deviations from the preregistered analysis plans are reported transparently. The stimuli used for this study can be found at: https://osf.io/hcrv6/files/osfstorage. The data used in the analyses reported below can be found at: https://osf.io/kyd6r/files/osfstorage.

Participants

Sample size was preregistered and selected based on prior work using the same experimental design and stimuli with younger adults (Silcox & Payne, 2021). We obtained consent for 48 community-dwelling older adults (34 women, mean age = 70.7 years, range = 60–85 years) who were paid $15 per hour for their participation. The sample was 2.1% Asian and 91.7% Caucasian, with 2.1% reporting more than one race and 4.2% electing not to disclose their race. In addition, 2.1% reported receiving at most a high school education, 52.1% completed at most some college, and 41.7% completed at most some postgraduate education, with 4.2% electing not to disclose their education experience. All participants were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) and reported being native speakers of English. All participants were within a normal range on the Montreal Cognitive Assessment (MoCA; mean = 27.2, range = 23–30; Carson, Leach, & Murphy, 2018; Waldron-Perrine & Axelrod, 2012) except for one participant who had a score of 21. We examined this individual's performance on the memory outcomes detailed below, and they did not appear to be an outlier in any way. Therefore, we decided to include this individual in the analyses.

Two hearing assessments, pure-tone audiometry and speech recognition thresholds (SRTs), were administered. Test signals were presented through an MA 41 audiometer via RadioEar IP-30 insert air-conduction earphones, and the modified Hughson-Westlake procedure was used to find the lowest level (in dB HL) at which the signal was perceived correctly at least 50% of the time. For pure-tone audiometry, thresholds were found at octave intervals from 250 to 8000 Hz and at 6000 Hz for each ear, plus inter-octaves (750, 1500, and/or 3000 Hz) if thresholds at adjacent octaves differed by 20 dB or more. SRTs were measured by having participant repeat recorded spondees (Central Institute for the Deaf; Auditec). Mean three-frequency (1000–4000 Hz), pure-tone average thresholds were 27.0 dB HL (range = 8.3–56.7) and 27.8 dB HL (range = 8.3–62.5), and mean SRTs were 25.2 dB HL (range = 10–55) and 26.1 dB HL (range = 5–60) for the right and left ears, respectively.

Materials

The stimuli used were identical to those used in Silcox and Payne (2021); details may be found there about how they were created, normed, and recorded. Stimuli included three types of sentences: high-constraint sentences with an expected sentence-final word, high-constraint sentences with an unexpected sentence-final word, and low-constraint sentences with an unexpected sentence-final word. Each high-constraint sentence context had two possible sentence-final words (either expected or unexpected); the low-constraint sentences used the same unexpected sentence-final word as the high constraint sentence. Examples of each stimulus type follow (target words are underlined):

(1) High-constraint context, expected target word: *The prisoners were planning their <u>escape</u>.*
(2) High-constraint context, unexpected target word: *The prisoners were planning their <u>party</u>.*
(3) Low-constraint context, unexpected target word: *Larry chose not to join the <u>party</u>.*

To create an acoustically challenging condition, power spectrum-matched noise was generated and added to each audio file. A SNR of +3 dB was chosen based on prior work showing that it increases experienced effort while maintaining high recognition performance (e.g., Crandell, Silcox, Ferguson, Lohani, & Payne, 2022; Payne et al., 2022; Silcox & Payne, 2021).

Procedure

The study followed a 2 (background: quiet or noise) × 3 (target word type: high-constraint expected, high-constraint unexpected, and low-constraint unexpected) design. There were 40 trials per condition for 240 trials. To ensure that each stimulus was used in each of the six experimental conditions, we created separate counterbalanced lists (see Silcox & Payne, 2021, for details on how the counterbalancing was achieved).

Before beginning the experimental tasks, participants were consented and completed a demographics questionnaire. Next, participants were administered the MoCA (Nasreddine et al., 2005), completed a short-form reading span task (see Oswald, McAbee, Redick, & Hambrick, 2015), and had their verbal ability tested using the extended range vocabulary test from ETS Kid of Factor Referenced Cognitive Tests (Ekstrom, Dermen, & Harman, 1976). The participants were then given hearing assessments as outlined above. Finally, the participants completed a short speech shadowing task to ensure that the level of background noise used during the experimental task was at a level that the speech could still be successfully perceived. For this speech shadowing task, participants heard three different test sentences that were recorded by the same speaker and used the same type and level of background noise that was used for the experimental sentences. These procedures were identical to those used in Silcox and Payne (2021), with one exception. Silcox and Payne (2021) did not administer the MoCA to their younger adult sample.

After performing these assessments, participants completed two tasks while seated 55 cm from a monitor in a chinrest to stabilize their heads and with ambient lighting of 140 lux: (1) a sentence listening task in which they heard 240 sentences (40 sentences in each of the six conditions described above) and (2) a combined sentence recognition and cued word recall memory test. For the sentence listening task, there was no online task to be completed. Participants were instructed that they would be completing a memory test after hearing the list of sentences. Each trial consisted of 2000 msec of silence, the sentence, and another 2000 msec of silence, all with a fixation cross presented in the middle of the screen for the duration of the trial. Stimuli were presented via the audiometer to their better hearing ear only as determined by their SRTs (in contrast, stimuli were presented binaurally in Silcox & Payne, 2021). To accommodate variation in hearing acuity, stimuli were presented at 40 dB above each participant's SRT with an upper limit of 70 dB HL to avoid discomfort and the potential for increased masking at very high presentation levels (Dubno, Horwitz, & Ahlstrom, 2005). For the memory test, participants were visually presented with 120 test sentence frames, each with the sentence-final word missing. They were to first mark whether they recognized a sentence as one that they had heard during the sentence listening task. For items that the participants reported remembering, they were tasked with recalling the word that they remembered hearing at the end of the sentence. Of the 120 test sentence frames that were presented, 60 were old items that had been heard previously (10 sentences from each of the six experimental conditions) and 60 were new items not previously presented. These 60 new sentence frames were semantic foils of previously heard (old) sentences, created by taking two to four meaning-bearing words from an old sentence that they had actually heard and creating new, semantically similar sentences. Foils were included to increase the difficulty of the recognition task so that participants were less likely to perform at ceiling (see Payne et al., 2022; Silcox & Payne, 2021; Koeritzer, Rogers, Van Engen, & Peelle, 2018).

EEG Recording and Processing

During the sentence listening task, we recorded EEG data from 32 evenly spaced silver–silver chloride actiCap slim active electrodes (Brain Products) following the standard International 10–20 localization system for 32 electrodes and kept electrode impedances below 20 kOhms. Online, electrodes were referenced to the TP10 electrode. Offline, electrodes were rereferenced to the average of the TP9 and TP10 electrodes, which are close to the left and right mastoids, respectively. Before analysis, we used independent components analysis to identify and remove eye blink artifacts from the EEG following the recommendations given by Luck (2022). Specifically, we copied the original EEG data—excluding the virtual channels—and, on the copy, we epoched the data −1000 to 2000 msec time locked to the onset of the target word. We then bandpass filtered the data from 1 to 30 Hz, down sampled to 100 Hz and ran independent components analysis using the infomax algorithm, implemented with the *runica* routine in EEGlab. We then used ICLabel (Pion-Tonachini, Kreutz-Delgado, & Makeig, 2019) to inspect each independent component and to identify which components contained predominantly ocular artifact-related variation. Once these components were identified, the independent component weights were transferred to the original data and these components were removed before transforming the data back into EEG sensor space (average number of components removed = 1.15, range = 1–2).

The resulting EEG data were then down-sampled to 250 Hz and bandpass filtered from 0.1 to 30 Hz. The data were then epoched 100 msec before and 1200 msec after the onset of the target word. The epoched data were then baseline corrected by subtracting the averaged amplitude within the −100- to 0-msec time window from the entire epoch. Epoched data were then examined for any remaining artifacts (i.e., flatlines, large signal drifts, residual eye movement artifacts). Any trials that were flagged as containing artifacts were excluded from analysis. Thresholds used for each artifact detection algorithm were selected for each individual participant through a condition-blind inspection of the data. Any participants that had greater than 40% of their data flagged as containing artifacts were removed from any subsequent analyses.

Pupillometry Recording and Processing

An Eyelink 1000 Plus desktop mounted infrared eye tracker camera (distributed by SR Research) was used to continuously measure pupil size from the right eye at a rate of 1000 Hz. Pupil size data were down sampled offline to 50 Hz and, for visualization purposes, epoched 200 msec before and 3000 msec after the onset of sentence audio. For analytic purposes, the data were epoched from 1000 msec before the onset of the sentence-final word to 0 msec before the onset of the sentence-final word and were baseline corrected using the mean pupil size −200 to 0 msec before the onset of the sentence. The epoched pupil data were then examined for artifacts including eye blinks and pupil dilation speed outliers (which can occur when the camera temporarily detects eyelashes or corrective lenses as part of the pupil and are seen as implausibly fast dilations of the pupil). Dilation speed outliers were removed, and blinks had 50 msec of data points removed on either side of the blink (blinks were defined as gaps in continuous data of more than 75 msec). Any trial missing 40% or more of the data points was flagged and excluded from future analyses. If a trial was not flagged, missing data points were filled in by linear interpolation and the interpolated data were run through a 10-Hz, low-pass Butterworth filter. Finally, each trial was baseline corrected by dividing each time point by the mean pupil size measured during the 200 msec before the onset of the sentence. This yielded the proportion change from baseline at each time point. This same baseline was used for both epochs described above.

Electrophysiological Data Analyses

There were two general classes of analyses on the ERP data. The first were mean amplitude analyses, and the second were onset latency analyses. For the mean amplitude analyses, we ran one set of a priori analyses on the N400 ERP response using the a priori window of 300–500 msec in six posterior electrodes (CP1, CP2, Cz, P3, P4, and Pz), where N400 effects are typically seen. We then ran two exploratory analyses on an apparent left-lateralized, late frontal ERP response using an exploratory window of 500–1200 msec in six frontal electrodes (Fp1, Fp2, F3, F4, F7, and F8). Separate models were created for the left (Fp1, F3, F7) and right (Fp2, F4, F8) frontal channels. All three of these mean amplitude analyses followed the same pattern. First, a linear mixed-effects model was fit with mean amplitude as the dependent variable and with Target Word Type (high constraint, expected vs. high constraint, unexpected vs. low constraint, unexpected), Background Noise (quiet vs. noise), and their interaction as independent variables. Random intercepts for participant and electrode were also included, which was the maximal random effects structure that allowed for convergence (Barr, Levy, Scheepers, & Tily, 2013). Likelihood ratio tests were run to assess main effects and interactions. If significant effects were present, then post hoc pairwise analyses were run on the estimated marginal means from the model to better understand these effects.

ERP latency analyses were run in two steps. First, difference waves of the expectancy effect were created for each electrode via pointwise subtraction of the participant-level ERP waveforms for the high constraint, unexpected, and high constraint, expected conditions separately for the quiet and noise conditions (Groppe, Urbach, & Kutas, 2011). Next, to assess any general latency changes in effects, we created raster plots by running a one-sample *t* test at each time point for each electrode using these participant-level difference waves. We used the false discovery rate procedure to control for multiple comparisons (Benjamini & Hochberg, 1995). Any of these *t* statistics that survived control for multiple comparison were plotted in our raster plots (red for negative effects, blue for positive effects). Thus, only those places in time and space that were significant after controlling for multiple comparisons were plotted. This allows one to easily see where in time there are significant differences between the high constraint, unexpected, and high constraint, expected conditions (i.e., the expectancy effect). If there were any significant differences in the expectancy effect in the posterior N400 or late frontal response regions, we used a jackknife procedure to measure 50% peak onset latency (Kiesel, Miller, Jolicoeur, & Brisson, 2008; Ulrich & Miller, 2001). Specifically, for the N400 expectancy effect, we used an a priori time window of 200–600 msec on an average of six posterior electrodes (i.e., CP1, CP2, Cz, P3, P4, and Pz) and low-pass filtered the EEG data at 10 Hz before creating participant-level difference waves for analysis. The 50% peak latency was then calculated for each jackknifed subsample, and a jackknife-corrected *t* test was conducted to compare the onset latencies between the quiet and noise conditions. A similar exploratory procedure was done for the late frontal response separately for the left (i.e., Fp1, F3, and F7) and right (i.e., Fp2, F4, and F8) frontal electrodes. The time window used for the late frontal response was 350–1200 msec.

Memory Data Analyses

We ran two sets of analyses on the memory data: one on the recognition hit rate and one set on the recall accuracy. Following Silcox and Payne (2021), we used the single-trial hit rate as the recognition outcome, because this is the only outcome that could be meaningfully used for the pupillometry coupling analyses (described below). In addition, because the recognition portion of the memory test used sentence frames without the sentence-final word present, we collapsed across the high-constraint expected and the high-constraint unexpected conditions. Thus, our recognition memory analysis compared the high- and low-constraint sentences. For this a priori analysis, we first built a generalized linear mixed-effects model, assuming a binomial distribution using a logit link function. We used single-trial hit or miss data as the outcome variable with sentence-level Contextual Constraint (high vs. low constraint), Background Noise (quiet vs. noise), and their interaction as predictor variables. A random intercept for participant, random slope for sentence context, random slope for noise, and a random intercept for item were included, which was the maximal random effects structure that allowed for convergence (Barr et al., 2013). We then ran a likelihood ratio test on the fixed effects to assess whether there were any main effects or an interaction. Odds ratios were then inspected to interpret the effect size magnitude of any significant effects.

For the recall data, single-trial accuracy data were used as the outcome variable for a generalized linear mixed-effects model, again assuming a binomial distribution using a logit link function. As predictors, we used sentence-final word type (high-constraint, expected vs. high-constraint, unexpected vs. low-constraint, unexpected), background noise (quiet vs. noise), and their interaction. For the random effects structure, we used a random intercept for participant with a random slope for noise and a random intercept for item as this was the maximal random effects structure that allowed for convergence (Barr et al., 2013). Likelihood ratio tests were used to assess whether there were any significant main effects or interaction on the fixed effects. Odds ratios were then inspected to interpret the effect size magnitude of any significant effects.

Pupillometry Data Analyses

Mean proportion change in pupil size from baseline was calculated 1000 to 0 msec before the onset of the sentence-final word. Participant-level means were then calculated separately for sentences presented in quiet and noise, and these means were used as the dependent variable in a linear mixed-effects model with Background Noise as the predictor and a random intercept for participant.

ERP-Pupillometry Coupling Data Analyses

To assess the relationship between the pupillary response and the N400 ERP response, an a priori analysis looking at the single-trial relationship between these two responses in noise was performed. For this analysis, a linear mixed-effects model was fit with single-trial N400 mean amplitude as the outcome variable with sentence-final word type (high constraint, expected vs. high constraint, unexpected vs. low constraint, unexpected), singe-trial pupil mean amplitude (participant-mean standardized), and their interaction as predictors. We also included a random intercept for participant and a random intercept for item. We then ran a likelihood ratio test and post hoc pairwise *t* tests on the estimated marginal means. As per our preregistration, this model was only run for the noise condition, as in Silcox and Payne (2021). However, we also ran an exploratory model that was identical, but for the data in the quiet condition. In addition, we ran analogous exploratory models for the late frontal ERP response.

We also ran a similar set of a priori latency analyses as above but rather than comparing the N400 50% mean amplitude onset latency between quiet and noise, we divided the ERP data in noise into trials in which there was either a larger or smaller pupillary response. This was done by calculating the median mean pupil size for each participant in noise and classifying trials with a mean pupil size below the median as "small" and those with a mean pupil size above the median as "large." We then ran a jackknife latency analysis identical to that described above comparing the onset latency for large and small pupil response trials. As an exploratory analysis, we likewise did this for the N400 response in quiet and for the late frontal response both in noise and quiet.

Memory-Pupillometry Coupling Data Analyses

For recognition memory outcomes, we fit a generalized linear mixed-effects model to the single trial data, assuming a binomial distribution using a logit link function. As the outcome variable, we used single-trial level recognition accuracy. As the predictors, we used sentence-level Contextual Constraint (high- vs. low-constraint), single-trial mean Proportion Change in Pupil Size (participant-mean standardized), and their interaction. We included a random intercept for participant. As per our preregistration, this model was only run on the data for the noise condition, as Silcox and Payne (2021) had done previously. However, we ran an identical exploratory analysis for the data in the quiet condition. In addition, we ran a similar a priori model for single-trial recall accuracy in noise. The only difference with this model is that it used single-trial recall accuracy as the outcome variable and replaced the sentence-level context predictor with a sentence-final Word Type Predictor (high-constraint, expected vs. high-constraint, unexpected vs. low-constraint, unexpected). An additional exploratory model on single-trial recall accuracy in quiet was also run.

RESULTS

ERP Mean Amplitude Results

Grand average ERPs across the left and right frontal electrodes and posterior electrodes can be found in Figure 1. Visual inspection of the ERP waveforms shows a typical and expected N400 context effect over posterior electrodes. However, we also observed a frontal response that appeared to distinguish expected target words from unexpected targets and also appeared sensitive to the noise manipulation. This was surprising given that Silcox and Payne (2021) found no evidence of a frontal response that was sensitive to context or noise in younger adults. Therefore, we decided to do exploratory analyses on this apparent frontal response in addition to the confirmatory analyses on the N400. For the a priori N400 mean amplitude analyses, our model revealed that there was a significant main effect of Target Word Type, $\chi^2(2) = 532.89, p <$ .001, and no significant main effect of Background Noise, $\chi^2(1) = 0.86, p = .35$. However, these main effects were qualified by a significant interaction between Background Noise and Target Word Type, $\chi^2(2) = 8.42, p = .015$. This interaction was explored by calculating estimated marginal mean pairwise contrasts. These contrasts are summarized

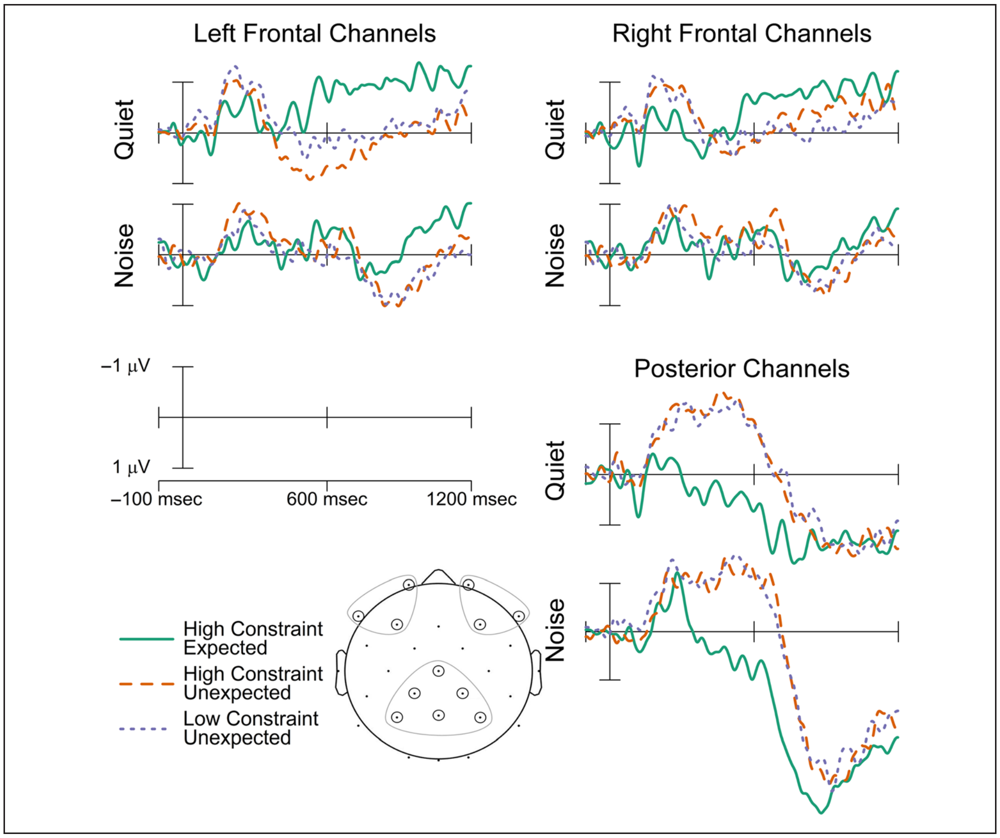


Figure 1. ERPs as a function of target word type and noise. These ERPs were averaged across the indicated electrodes.

Table 1. Pairwise Post Hoc Contrasts for Two of the Univariate ERP Mean Amplitude Models

N400 Mean Amplitude Model Post Hoc Contrasts			
Contrast	*Dif. Est.*	*z*	*p Value*
Quiet: high, expected vs. high, unexpected	1.72	16.48	<.001
Noise: high, expected vs. high, unexpected	1.31	12.52	<.001
Quiet: high, expected vs. low, unexpected	1.59	15.25	<.001
Noise: high, expected vs. low, unexpected	1.30	12.37	<.001
Quiet: high, unexpected vs. low, unexpected	−0.13	−1.20	.28
Noise: high, unexpected vs. low, unexpected	−0.02	−0.16	.87
High, expected: quiet vs. noise	0.18	1.75	.11
High, unexpected: quiet vs. noise	−0.23	−2.21	.04
Low, unexpected: quiet vs. noise	−0.12	−1.16	.28
Late Left Frontal Response Model Post Hoc Contrasts			
Quiet: high, expected vs. high, unexpected	−1.19	−6.05	<.001
Noise: high, expected vs. high, unexpected	−0.47	−2.40	.03
Quiet: high, expected vs. low, unexpected	−0.92	−4.67	<.001
Noise: high, expected vs. low, unexpected	−0.59	−3.02	<.001
Quiet: high, unexpected vs. low, unexpected	0.27	1.38	.25
Noise: high, unexpected vs. low, unexpected	−0.12	−0.63	.61
High, expected: quiet vs. noise	−0.69	−3.49	<.001
High, unexpected: quiet vs. noise	0.03	0.17	.87
Low, unexpected: quiet vs. noise	−0.36	−1.84	.11

The top panel reports the N400 response. The bottom panel reports the contrasts for the late frontal response in the left frontal electrodes. All contrasts were controlled for multiple comparisons.

in Table 1. Confirming visual inspection, we observed a classic N400 pattern, with a generally larger N400 mean amplitude for unexpected sentence-final words than for expected sentence-final words. The mean amplitudes for high constraint, expected words, and low constraint, unexpected words were generally unaffected by the noise manipulations. However, we found a significant reduction of the mean amplitude in noise for high constraint, unexpected sentence-final words, compared with those words heard in quiet.

For the exploratory analysis on the left late frontal response, we found that there were significant main effects of Target Word Type, $\chi^2(2) = 43.64, p < .001$, and of Noise, $\chi^2(1) = 8.88, p = .0029$. However, these main effects were qualified by an interaction between Target Word Type and Noise, $\chi^2(2) = 6.70, p = .035$. We explored this interaction by calculating post hoc pairwise contrasts that can be found in Table 1. In general, we found that the response to high constraint, expected words was significantly more negative in both quiet and noise than to unexpected words. However, this response to high constraint, expected words was significantly and selectively reduced in noise. Likewise, we saw that the response to unexpected words did not differ as a function of constraint, nor was the response to unexpected words affected by the presence of noise (regardless of sentence constraint).

This same exploratory model was run in the right frontal electrodes. We found that there were significant main effects of Target Word Type, $\chi^2(2) = 11.61, p = .0030$, and Background Noise, $\chi^2(1) = 9.14, p = .0025$. These effects were not qualified by an interaction, $\chi^2(2) = 3.33, p = .19$. We found that there was a general shift of the late frontal response to be more positive in background noise (estimated difference = 0.33, $z = 3.02$, $p = .0025$). We also found that the late right frontal response was generally more negative to high constraint, expected words than to high constraint, unexpected words (estimated difference = −0.31, $z = -2.27, p = .023$) and to low constraint, unexpected words (estimated difference = −0.45, $z = -3.34, p < .001$). The response to high constraint, unexpected words was not significantly different than to low constraint, unexpected words (estimated difference = -0.14, $z = -1.06$, $p = .29$).

ERP Latency Results

Figure 2 displays the ERP expectancy effect (i.e., high constraint, unexpected minus high constraint, expected) in both quiet and in background noise. Note that from Figure 2B and C that the expectancy effect shows a canonical N400 centro-posterior distribution. There is also a notable expectancy effect that is later in time and positive in polarity that is frontally distributed, which is lateralized slightly to the left on the scalp. The ERP difference waves (Figure 2A), the raster plots (Figure 2B), and the scalp topographies (Figure 2C) all show an apparent delay in the onset latency of both the N400 and late frontal expectancy effects. This was confirmed by the a priori jackknife analysis on the N400 expectancy effect and the exploratory jackknife analyses on the left and right frontal expectancy effects. We found that the average 50% peak onset latency of the N400 expectancy effect was 265.96 msec ($SE = 0.58$) in quiet and 345.83 msec ($SE = 0.38$) in noise (a difference of 79.58 msec). The difference in onset latencies was found to be statistically significant, t_corrected(47) = -2.20, $p = .033$. For the late positive expectancy effect measured in the left frontal electrodes (i.e., Fp1, F3, F7), we found that the average onset latency of the effect was 437.96 msec ($SE = 0.32$) in quiet and 839.58 msec ($SE = 0.56$) in noise (a difference of 401.63 msec). This delay was statistically significant, t_corrected(47) = -12.46, $p < .001$. For the right frontal electrodes (i.e., Fp2, F4, F8), we found that the expectancy effect in quiet had an average 50% peak onset latency of 517.63 msec ($SE = 1.16$) and in noise the onset was 854.13 msec ($SE = 0.60$; a difference of 336.50 msec). This delay in the late right frontal expectancy effect was statistically significant, t_corrected(47) = -4.96, $p < .001$.

To assess whether there were any constraint effects (high constraint, unexpected minus low constraint,

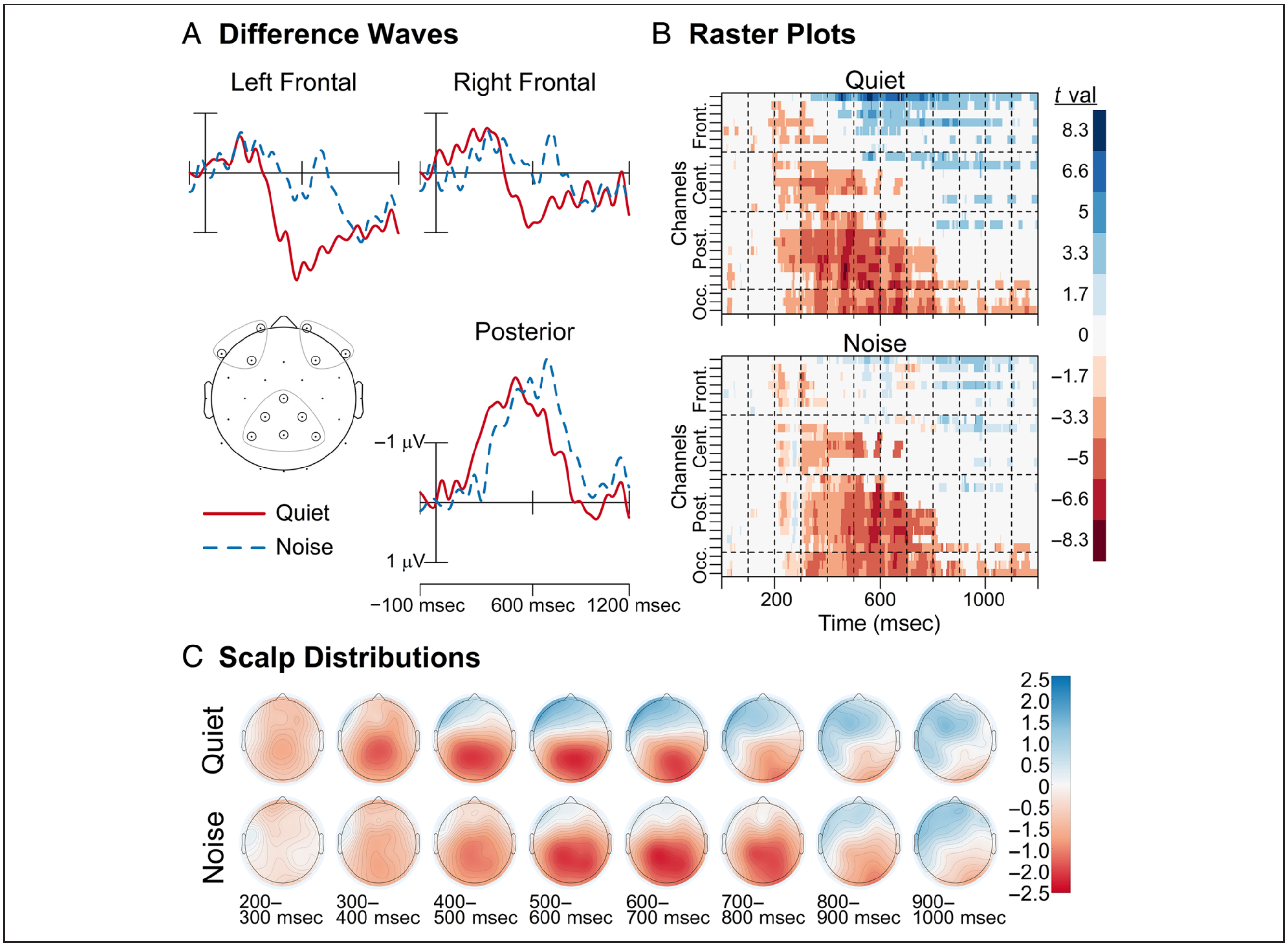


Figure 2. The expectancy effect is displayed as a function of background noise. The expectancy effect is found by taking pointwise difference between unexpected and expected words in high constraint sentence contexts (i.e., high constraint, unexpected minus high constraint, expected). (A) The difference wave ERPs for sentences heard in quiet (red, solid line) and in noise (blue, dashed line). The jackknife analyses for the N400 expectancy effect were conducted in the 200- to 600-msec time window. The jackknife analyses for the late frontal expectancy effect were conducted in the 350- to 1200-msec time window. (B) The false discovery rate corrected raster plots of the expectancy effect in quiet and in noise. See text for details on how the raster plots were constructed. (C) The scalp topography maps (in 100-msec bins) from 200 msec to 1000 msec. The top row is the topography of the quiet condition, and the bottom row is the topography in the noise condition.

unexpected), we built raster plots that revealed no significant differences between these conditions at any time point for any electrode. Indeed, Figure 1 shows that—at least for the frontal and posterior electrodes—the ERPs for the high constraint, unexpected condition, and for the low constraint, unexpected condition laid on top of each other. Taking this all together, we did not pursue any latency analyses for the constraint effect in the ERPs.

Memory Results

Figure 3A shows the results from the recognition portion of the memory test. For the a priori analysis on recognition memory hit rate, we found a significant main effect of Sentence Context, $\chi^2(1) = 34.30, p < .001$, such that the odds of recognizing a high constraint sentence were 2.94 times the odds of recognizing a low constraint sentence ($z = 6.16$, $p < .001$). However, there was no main effect of Background Noise, $\chi^2(1) = 1.31, p = .25$, nor a significant interaction between Sentence Constraint and Background Noise, $\chi^2(1) = 0.00, p = .95$.

Figure 3B displays the results of the recall portion of the memory test. Our a priori model of recall performance found that there was a significant main effect of Target Word Type, $\chi^2(2) = 189.62, p < .001$, but no significant main effect of Background Noise, $\chi^2(1) = 3.57, p = .059$. There was a significant interaction between Target Word Type and Background Noise, $\chi^2(2) = 6.20, p = .045$. The pairwise contrasts for post hoc comparisons on memory can be found in Table 2. These contrasts revealed that high constraint, expected words were generally better remembered than high constraint, unexpected, and low constraint, unexpected words. This was true for words heard in both quiet and in background noise. High constraint, expected, and low constraint, unexpected words were not significantly impacted by background

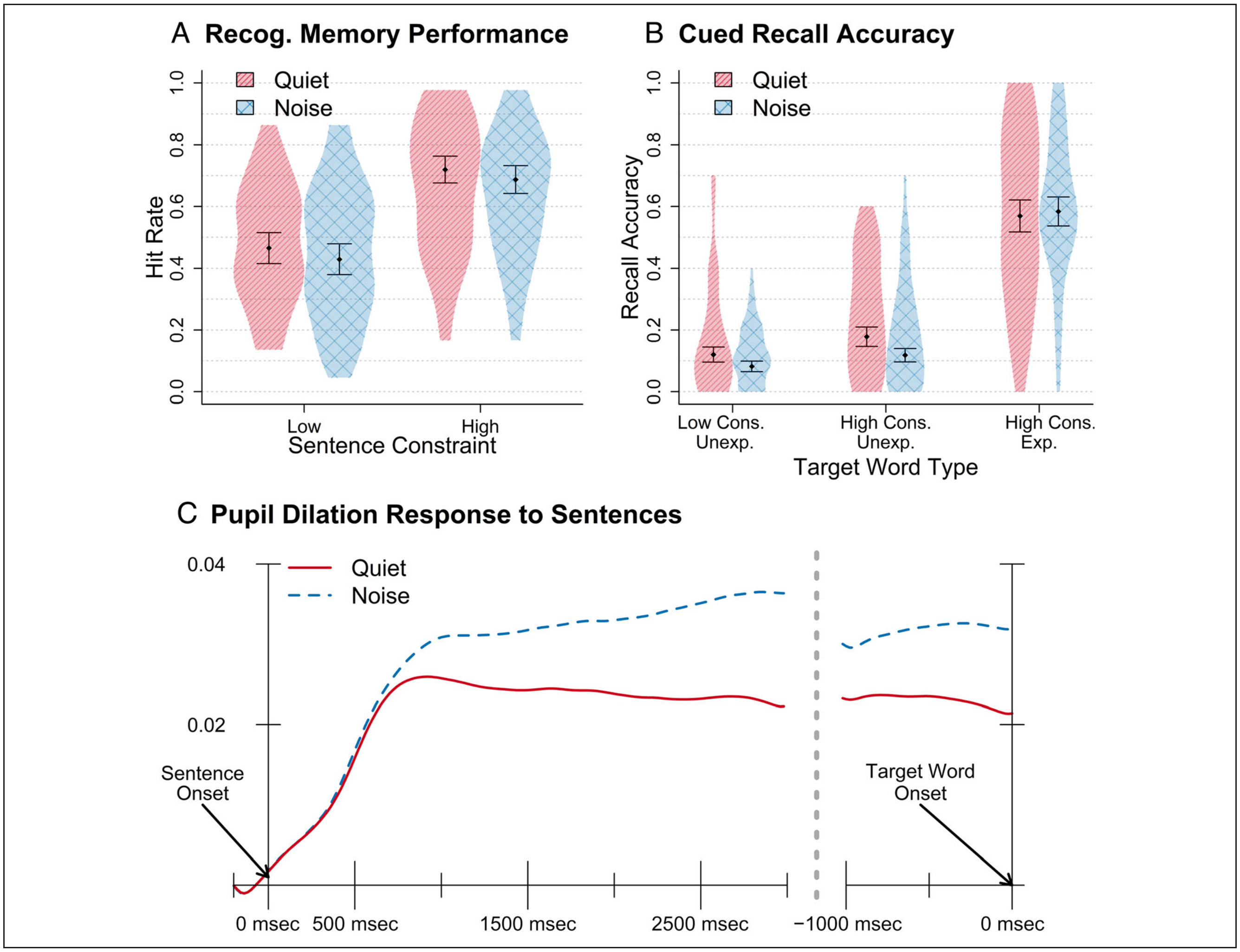


Figure 3. Behavioral and pupillary outcomes. (A) Results from the recognition portion of the memory test as a function of background noise and sentence constraint. Points show the estimated marginal means, and the error bars show the estimated standard error of those means. Shaded regions show the densities of the underlying distributions. The left-most column shows the hit rate for recognition memory. (B) Results from the recall portion of the memory test as a function of background noise and target word type. Points show the estimated marginal means, and the error bars display the standard error of those means. The shaded regions display the density of the underlying distributions. (C) The pupillary response was time locked to the sentence onset. The pupil response just before the target word onset is also displayed. Cons. = constraint; Exp. = expected; Unexp. = unexpected.

Table 2. Pairwise Post Hoc Contrasts for the Univariate Model on Recall Accuracy

Recall Accuracy Model Post Hoc Contrasts			
Contrast	*Est. OR*	*z*	*p Value*
Quiet: high, expected vs. high, unexpected	6.10	9.06	<.001
Noise: high, expected vs. high, unexpected	10.47	11.36	<.001
Quiet: high, expected vs. low, unexpected	9.67	9.81	<.001
Noise: high, expected vs. low, unexpected	15.77	11.36	<.001
Quiet: high, unexpected vs. low, unexpected	1.59	2.09	<.05
Noise: high, unexpected vs. low, unexpected	1.51	1.73	.10
High, expected: quiet vs. noise	0.94	−0.33	.79
High, unexpected: quiet vs. noise	1.62	2.34	<.05
Low, unexpected: quiet vs. noise	1.54	1.86	.08

OR = odds ratio.

noise but high constraint, unexpected words were significantly better recalled than low constraint, unexpected words in quiet but not in noise.

Pupillometry Analyses

The task evoked pupillary response is shown in Figure 3C. This figure displays the pupillary response time locked to the onset of the sentence as well as the change in pupil size relative to baseline just before the onset of the target word. A priori analysis on the mean amplitude 1000 msec before the onset of the sentence-final word revealed a main effect of background noise, $\chi^2(1) = 14.42$, $p < .001$, such that the pupillary response in noise was larger than the response in quiet, mean in quiet = 0.023, *SE* in quiet = 0.0043; mean in noise = 0.032, *SE* in noise = 0.0045; $t(49) = 4.06$, $p < .001$.

ERP-Pupillometry Coupling Analyses

Figure 4 shows the difference waves of the ERP expectancy effect (high constraint, unexpected minus high constraint, expected) separately for trials with smaller (below the intraparticipant median) versus larger (above the intraparticipant median) pupillary responses. The a priori model assessing the relationship between the N400 mean amplitude and the pupillary response found that there was a significant main effect of Target Word Type, $\chi^2(2) = 40.49$, $p < .001$, consistent with the model reported above in which pupil size was not used as a predictor. We also found a significant main effect of Mean Pupil Size, $\chi^2(1) = 3.84$, $p = .050$, such that with each 1 *SD* increase in pupil size the N400 mean amplitude increased (i.e., became more positive) by 0.19 uV ($z = 1.96$, $p = .050$). However, there was no significant interaction effect, $\chi^2(2) = 0.58$, $p = .75$. Reasoning that variation in pupil-mediated arousal/effort may impact speech processing even in less demanding

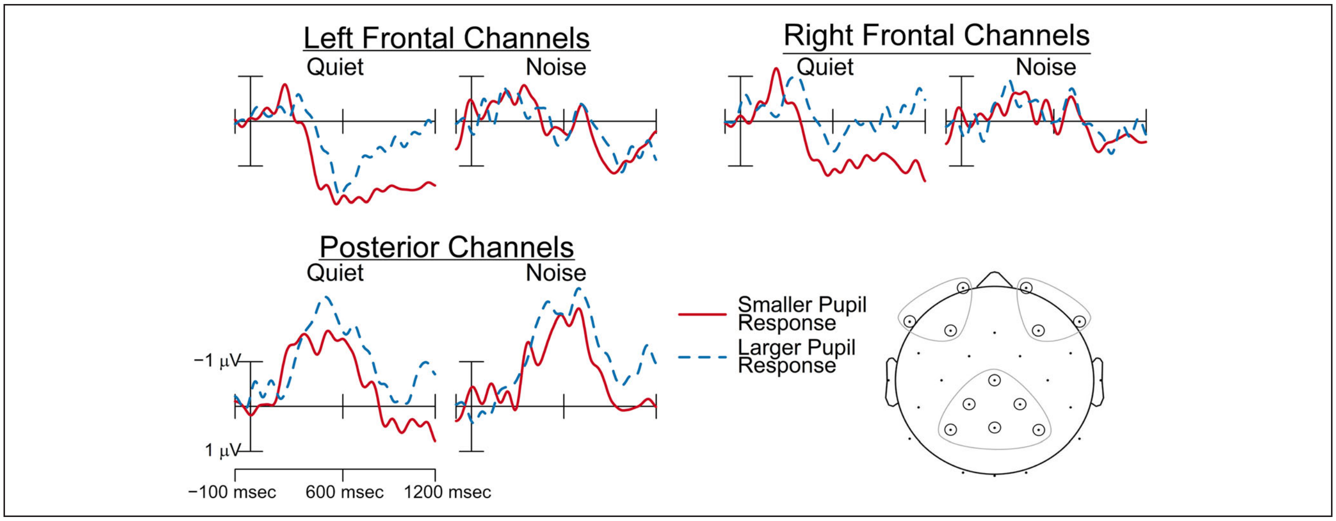


Figure 4. The ERP expectancy effect (high constraint, unexpected minus high constraint, expected) as a function of pupillary response. Difference waves reflect the expectancy effect, which is the high constraint, unexpected response minus the high constraint, expected response.

quiet trials for older adults, we also ran an exploratory analysis of the relationship between the pupillary response and the mean amplitudes of the N400 and late frontal responses in quiet. However, these exploratory models revealed no significant relationship between mean ERP amplitude and pupil size.

An a priori jackknife analysis was conducted to determine the relationship between the pupillary response and the 50% peak onset latency of the N400 expectancy effect in noise. This analysis showed that the average onset latency for trials in which there was a smaller pupillary response was 336.38 msec (*SE* = 0.28) and for trials with larger pupillary responses it was 378.71 msec (*SE* = 0.55). This is a difference of approximately 42.33 msec, which was not statistically significant, $t\_corrected(47) = -1.32$, $p = .19$. However, our exploratory latency analysis of this same effect in quiet showed that the average onset latency in quiet for the small pupil size trials was 207.67 msec (*SE* = 0.28) and was 302.21 msec (*SE* = 0.44) for larger pupil size trials (a difference of 94.54 msec). We found that this difference in onset latency was statistically significant, $t\_corrected(47) = -4.41$, $p < .001$. Our exploratory jackknife analysis of the late left frontal response in noise revealed an onset latency of 815 msec (*SE* = 0.61) for smaller pupil size trials and 700.04 msec (*SE* = 26.30) for larger pupil size trials (a difference of 177.38 msec), which was not statistically significant, $t\_corrected(47) = 0.09$, $p = .93$. Note the large standard error for the large pupil size trials here. It is likely that there was no frontal response here, leading to a highly unstable calculation of onset latency. Indeed, when inspecting Figure 4, one can see little evidence of a frontal response in noise for larger pupil trials. For this same analysis on the left frontal electrodes in quiet, we found an average onset latency of 409.38 msec (*SE* = 0.38) for smaller pupil size trials and an onset of 526.75 msec (*SE* = 0.31) for larger pupil size trials. This difference of 117.38 msec was statistically significant, $t\_corrected(47) = -5.36$, $p < .001$. For right frontal electrodes we found that, in noise, the onset latency was 694.75 msec (*SE* = 15.62) for smaller pupil size trials and 933.54 msec (*SE* = 0.81) for larger pupil size trials (a difference of 238.79 msec), which was not statistically significant, $t\_corrected(47) = -0.32$, $p = .75$. However, for quiet trials, the onset latency was 443.71 msec (*SE* = 0.78) for trials with a smaller pupillary response and 554.80 msec (*SE* = 0.81) for trials with a larger pupillary response (a difference of 112.08 msec), which was statistically significant, $t\_corrected(47) = -2.08$, $p = .043$. In summary, in noise, we did not find that the onset latency of ERP responses changed as a function of pupil size. However, in quiet, for both the N400 and the late frontal response, we found that a larger pupil size led to a delay in onset latency.

Figure 5 displays the raster plots of the expectancy effect (high constraint, unexpected minus high constraint, expected) both in quiet and in noise separately for trials

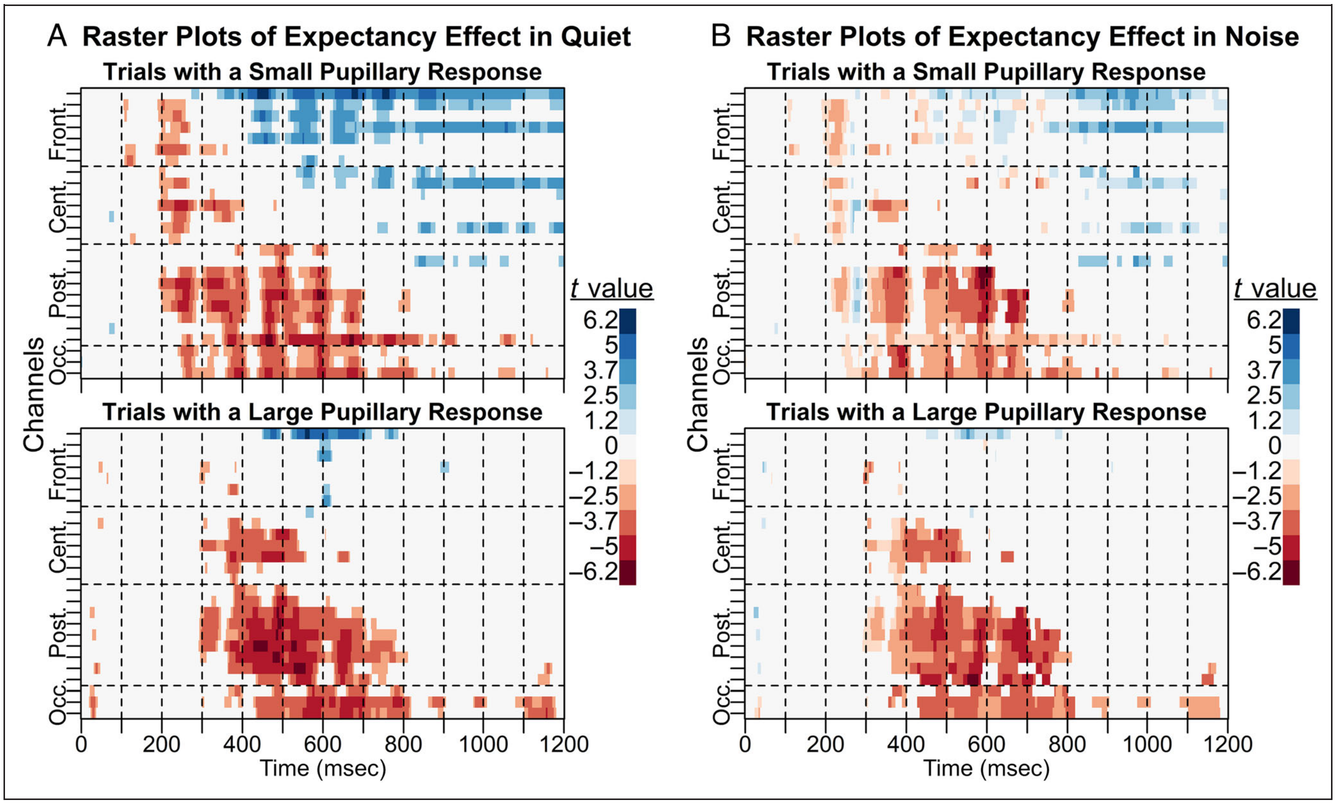
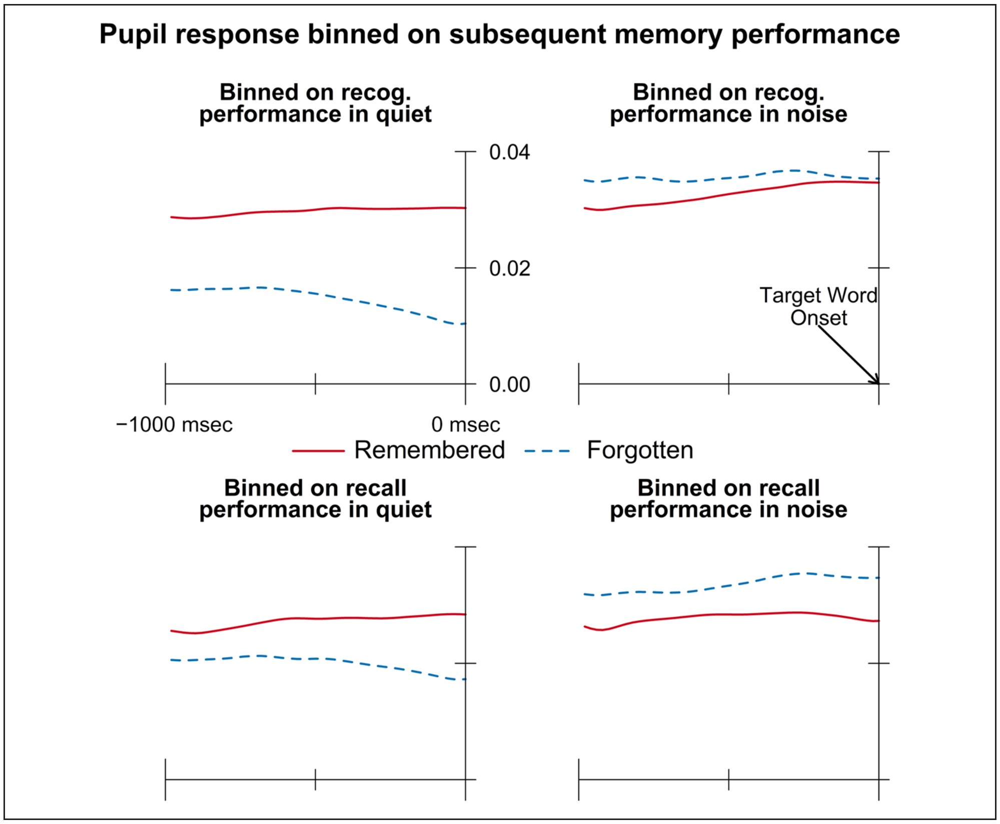


Figure 5. Raster plots of the expectancy effect (high constraint, unexpected minus high constraint, expected) that have been binned based on intrasubject median mean pupil sizes. (A) Raster plots for this effect in the quiet condition. (B) Raster plots for this effect in the noise condition.

Figure 6. The relationship between the pupillary response and subsequent memory performance. The pupillary response is binned on subsequent memory performance. The solid red line represents pupil response for trials that were subsequently remembered, and the dashed blue line is for trials that were subsequently forgotten.



with smaller (below the intraparticipant median) versus larger (above the intraparticipant median) pupillary responses. In quiet, one can see that there is a clear delay in the onset latency of the N400 expectancy effect for trials with larger pupil sizes compared with trials with smaller pupil sizes. This change in latency is less clear in noise as the N400 effect appears to be much less stable in noise for small pupil size trials, consistent with our jackknife latency analyses reported above.

The raster plot for the expectancy effect in quiet shows a clear decrease in the spatial and temporal extent of the frontal response for trials in which there was a larger pupillary response, in contrast to our ROI-based mean amplitude analyses. The frontal expectancy effect is likewise much less stable in noise, but there is evidence to suggest that it is present for trials with smaller pupil sizes and almost completely absent for trials with larger pupil sizes.

Memory-Pupillometry Coupling Analyses

Figure 6 shows the pupillary response before the target word binned on whether a sentence or word was subsequently forgotten or correctly remembered. For the a priori model for recognition memory performance in noise, we found no significant main effect of Pupil Size, $\chi^2(1) = 3.60, p = .06$, and no significant interaction, $\chi^2(1) = 0.49, p = .48$. However, for exploratory analyses of the quiet trials, we found a significant main effect of Pupil Size, $\chi^2(1) = 10.24, p < .001$, such that for each 1 *SD* increase in pupil size, the odds of recognizing a sentence heard in quiet increased by 1.18 times ($z = 2.13, p = .033$). There was no significant interaction, $\chi^2(1) = 0.30, p = .59$.

Our a priori model looking at the relationship between recall accuracy and the pupillary response in noise found a significant main effect of Pupil Size, $\chi^2(1) = 5.01, p < .05$, such that for each 1 *SD* increase in pupil size, the odds of recalling a target word heard in noise decreased by 1.25 times ($z = -2.24, p = .025$). There was no significant interaction effect, $\chi^2(2) = 5.10, p = .08$. An exploratory analysis for quiet trials yielded no significant main effect of Pupil Size, $\chi^2(1) = 0.49, p = .48$, and no significant interaction effect, $\chi^2(2) = 0.39, p = .82$.

DISCUSSION

In the current study, we investigated how context processing in older adulthood is modulated by acoustic challenge using simultaneous EEG-pupillometry and subsequent memory. Most current models of listening effort suggest that even moderate increases in acoustic challenge could lead to disruptions in the way that speech is processed and remembered (e.g., Rogers & Peelle, 2022; Pichora-Fuller et al., 2016). To the best of our knowledge, this is the first study to investigate the effects of acoustic challenge and listening effort on ERPs in older adults. Overall, for average behavioral, neural, and physiological responses, we generally replicated in older adults what has been seen previously in ERP responses for younger adults (Silcox & Payne, 2021). That is, we observed delays in onset latency and reductions in amplitude in ERP responses and increases in pupil dilation in noise relative to quiet. Likewise, we saw that older adults were negatively impacted in their memory for unexpected words (i.e., words that were not predictable given the preceding context) heard in noise relative to quiet, but this noise effect was absent for expected words (i.e., words that were predictable given the preceding context). This replicates what has been found previously in cognitive audiology research. For example, McCoy and colleagues (2005) reported that older adults had difficulty in recalling words heard in background noise if the words heard were not predictable based on the preceding context. If the words fit well semantically with the preceding context, then the older adults were much more likely to recall these words despite the presence of background noise (for a review of similar studies, see Payne & Silcox, 2019).

Beyond average responses, however, what was novel about the current study is that we investigated—at the single trial level—the relationship between pupillary responses (as a marker of arousal-linked changes in effort allocation) and ERP and memory outcomes. We found a complex but clear relationship between pupil-linked arousal and these outcomes. In noise, we found that on trials where older adults increased their effort-related arousal response (measured by pupil size), there was: (1) a decrease in memory (significantly for recall and marginally for recognition (2) a decrease in the spatial and temporal extent of late frontal ERP responses (as seen in the mass univariate analyses in Figure 5), but (3) no evidence for a change in the N400 response. Overall, these findings suggest that an increase in effort-related arousal associated with acoustic challenge did not bring any noticeable benefits but was associated with significant costs for the aging listener. On the other hand, when examining these same relationships in the less demanding quiet trials, we found that increases in pupil size were associated with better subsequent recognition memory, but that this came at the cost of delays in both the N400 and late frontal expectancy effects and decreases in the spatial and temporal extent of the late frontal response. Collectively, these findings suggest that with increased acoustic challenge, increases in the effort-related arousal do not bring about any apparent benefits but only costs to older listeners, whereas in less demanding listening environments, increases in arousal can lead to the benefit of better memory at the cost of real-time, lexico-semantic processing during word recognition. In the following sections, we discuss these findings in more detail as well as their implications for models of listening effort and semantic processing in aging.

The Impact of Acoustic Challenge on Language-related ERPs

Both in quiet and in noise, we saw that older listeners showed robust N400 expectancy effects—larger N400 amplitudes to words that were unexpected compared with expected based on their prior context, consistent with prior results showing N400 context effects in older adults (Wlotko & Federmeier, 2012; see review in Payne & Silcox, 2019). Although the N400 response has been well documented in older adults, the impact of acoustic challenge on older adults' N400 response is not as well understood. There is a small but growing body of research that has found that the N400 tends to be reduced in amplitude and delayed in its onset latency when listening to acoustically challenging speech (e.g., Hsin, Chao, & Lee, 2023; Wambacq et al., 2023; Silcox & Payne, 2021; Kyong, Kwak, Han, Suh, & Kim, 2020; Obleser & Kotz, 2011; Aydelott, Dick, & Mills, 2006). However, to the best of our knowledge, this work has all been conducted in young normal-hearing listeners. The results from the current study show that acoustic challenge also has an impact on the N400 in older listeners, with a reduction in the amplitude of the N400 expectancy effect as well as a delay in the onset latency in noise compared with quiet. These findings suggest that, like the young (Silcox & Payne, 2021), older adults show an overall reduction in the efficiency of semantic retrieval when listening to acoustically challenging speech.

We also found evidence for a frontally distributed, left lateralized ERP response that appeared to distinguish between the high-constraint expected condition and the unexpected conditions (regardless of constraint). Late frontal ERP responses, typically occurring after the onset of the N400, have been consistently observed across multiple studies (e.g., Hubbard & Federmeier, 2021; Kuperberg, Brothers, & Wlotko, 2020; Payne & Federmeier, 2017; DeLong, Quante, & Kutas, 2014; Wlotko & Federmeier, 2012; Federmeier, Kutas, & Schul, 2010; Federmeier, Wlotko, De Ochoa-Dewalk, & Kutas, 2007). However, these later frontal responses are not always observed (e.g., Silcox et al., 2023; Stone, Nicenboim, Vasishth, & Rösler, 2023; Silcox & Payne, 2021; Stites, Payne, & Federmeier, 2017). Moreover, even when observed, features of the study design and stimuli can make it difficult to differentiate between two distinct but spatiotemporally overlapping late frontal responses. Federmeier and colleagues (2007) first characterized a late anterior positivity observed to words likely to elicit a *prediction violation*, that is, an unexpected but plausible word within a context constraining toward a different word. However, a similarly distributed late frontal negativity can oftentimes also be observed to expected words within a moderate-to-highly constraining context (Lai, Payne, & Federmeier, 2024; Ng, Payne, Steen, Stine-Morrow, & Federmeier, 2017; Payne & Federmeier, 2017). Although in some studies, it can be difficult to differentiate between the two frontal responses (e.g., Brothers, Swaab, & Traxler, 2017; Delong, Urbach, Groppe, & Kutas, 2011), Lai and colleagues (2024) have recently shown that the two responses are dissociable in their response to task demands. For example, whereas the anterior positivity to prediction violations is not impacted by explicit instructions to engage in prediction, the frontal negativity to expected words is strongly modulated by task demands. Similarly, we observed that effects of noise appeared to modulate the frontal response selectively for expected words in highly constraining contexts, such that it was reduced in magnitude and delayed in latency in noise compared with quiet, suggesting that the neurocognitive processes signaled by the frontal negativity were impaired in acoustically challenging environments.

The functional significance of the frontal negativity to expected words in constraining contexts is still not very well understood. However, the recent findings from Lai and colleagues (2024) may shed light on the processes indexed by the frontal negativity. They found that the frontal negativity was specific to expected words in highly constraining contexts but only for words in which the participant judged the word as being an exact match to what they were expecting to see. Lai and colleagues therefore concluded that highly constraining sentences could still afford multiple possible interpretations that the participant maintains in parallel over the course of the sentence and so "the frontal negativity might then reflect processes associated with selecting and solidifying the message-level representation supported by encountering the most globally probable word" (p. 16). Much work still remains to be done to better understand the frontal negativity's functional role, but evidence from the current study suggests that acoustic challenge may diminish the capacity for older adults to engage in such high-level comprehension processes or may represent a strategic shift in the allocation of these frontal neural resources when encountering acoustically challenging speech, a point we return to below.

The Impact of Acoustic Challenge on Memory

In the current study, we observed several replications of the findings from Silcox and Payne (2021) in younger adults detailing how prediction-related processes can facilitate memory encoding and maintenance. We found that high-constraint sentences were generally remembered better than low-constraint sentences and that expected sentence-final words were recalled better than unexpected sentence-final words. In general, this adds to the growing literature that engaging in predictive processes can benefit memory (e.g., Silcox et al., 2023; Silcox & Payne, 2021; Gordon-Salant & Fitzgibbons, 1997; Kutas, 1993). The results from the current study also showed that prediction is not just helpful when a prediction turns out to be correct but also when it is incorrect. In quiet, we saw that unexpected words that were heard in highly constraining contexts were recalled with higher accuracy than unexpected words that were heard in low-constraint contexts. Importantly, these two types of unexpected words were controlled on cloze probability, meaning that they were equally unexpected. However, highly constraining contexts afford greater opportunity to engage in predictive processing, whereas this would be extremely difficult to do for a low constraint sentence. A recent study by Hubbard and Federmeier (2024) similarly found that unexpected words in strongly constraining contexts were recognized better than unexpected words in low-constraint contexts (see also Lai, Rommers, & Federmeier, 2021). Collectively, the results from the current study, and from Hubbard and Federmeier (2024), suggest that unexpected words may be processed and remembered differently depending on whether the preceding context supports prediction or not. Altogether, our findings suggest multiple routes through which older adults can benefit from context-driven predictions in benefiting subsequent memory in acoustically ideal environments.

Although we generally found similar memory results in older adults compared with the younger adults in Silcox and Payne (2021), the impact of acoustic challenge on memory was quite different. In young adults, Silcox and Payne (2021) saw that noise negatively impacted recognition of low-constraint sentences but not high-constraint sentences. Likewise, younger adults had a general reduction in recall for all sentence-final word types in noise. Silcox and Payne interpreted this as evidence that surface-level lexical representations and gist-based sentence-level representations were being differentially impacted by acoustic challenge. They argued that supportive sentential context may be more beneficial to sentence-level representations in helping to buffer against acoustic challenge but may be less beneficial for lexical representations. However, for older adults, there was no apparent impact of acoustic challenge on sentence recognition performance. For recall, however, we did see that the constraint benefit on memory for unexpected words was eliminated in noise. Thus, in the current study, the impact of noise on older adults' memory appeared to be specific to the benefit of implicit learning that occurs when encountering prediction errors.

Pupil-mediated Arousal and Context Processing

Although multiple models of listening effort have been proposed in recent years, at some level of abstraction, they all share some overlap in at least three key components—(1) the capacity required to successfully complete a task, (2) the capacity available for a person to complete a task, and (3) the capacity that is actually attributed to the task (Rogers & Peelle, 2022; Francis & Love, 2019; Ayasse & Wingfield, 2018; Peelle, 2018; Shenhav et al., 2017; Pichora-Fuller et al., 2016; Kurzban, Duckworth, Kable, & Myers, 2013). Effort is generally conceptualized as the actual response to the difficulty of a task. That is, effort does not directly reflect how demanding a task is but how an individual deploys additional cognitive and neural resources in response to that difficulty above and beyond the "default" resources required within an idealized context to compensate for increases in task demands.

Within this conceptualization, the pupillary response tracks well as a marker of (at least part of) the effort response. There are clear increases in the pupil response with increasing task demands, including in the current study. However, the pupil response typically only increases up to a certain point as task difficulty increases, after which the average pupillary response begins to decrease (e.g., Wendt et al., 2018). This is exactly what we might expect to see from a measure of effort. That is, effort would increase up to the point at which the task becomes too difficult to complete, after which the amount of effort exerted would begin to decrease. In other words, the pupillary response does not just reflect task demands but appears to co-vary with the amount of effort exerted. It is because the pupil responds in a way that mirrors an effort-related response that many cognitive hearing scientists in the past have argued that the pupil response indexes variation in effort (e.g., Silcox & Payne, 2021; Zekveld et al., 2010, 2018; Miles et al., 2017; Wendt et al., 2017; Winn et al., 2015; Zekveld & Kramer, 2014). In the remainder of the discussion, we will focus on three key pupillary findings from the current study: the average change in pupil size from quiet to noise, the relationships between trial-to-trial changes in pupil size and ERP/memory outcomes in noise, and, finally, these same relationships in quiet.

In the current study, we replicated the results of dozens of prior studies finding that there is generally an increase in pupil size with an increase in acoustic challenge (for a review, see Zekveld et al., 2018), including what Silcox and Payne (2021) found with younger adults listening to the exact same stimuli and performing the same task. Generally, our results show that the pupillary response (both on average and at the single-trial level) can be measured reliably despite senile miosis—whereas some have suggested that the pupillary response is less useful in older adults (Van Gerven, Paas, Van Merriënboer, & Schmidt, 2004). Critically, these findings complement a small but growing literature showing that reliable effects of acoustic challenge on the pupillary response can be observed in older adults (e.g., McGarrigle et al., 2021; Kuchinsky et al., 2013; Piquado, Isaacowitz, & Wingfield, 2010), bolstering its use as a measure in older adult listening research. The average change in the pupil's response to acoustic challenge has been argued to reflect a general change in transient arousal responses, which may reflect increases in LC-NE activity, that are part of the more general effort-related response (de Gee et al., 2017; Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Aston-Jones & Cohen, 2005). This arousal-related response may reflect the recruitment of additional cognitive and neural resources beyond the traditional perisylvian language network (see Rogers & Peelle, 2022; Peelle, 2018) to help compensate for the increase in acoustic challenge.

In addition to looking at the general change in pupil size from quiet to noise, we also examined the trial-to-trial covariation between the pupillary response and both ERP and memory outcomes as in Silcox and Payne (2021). In the current study with older adults, we found no evidence for a relationship between pupil size and the N400 response in noise. However, we did observe a significant relationship between the pupillary response and sentence-final word recall, such that trials with larger pupillary responses were associated with poorer word recall, consistent with Silcox and Payne (2021), as well as a similar but marginally significant effect for recognition memory performance. Moreover, we observed that the spatial and temporal extent of the late frontal expectancy effect (see Figure 5B) is decreased with increasing pupil size as well.

Silcox and Payne (2021) found that younger adults were able to recover their noise-impaired N400 response on trials with larger pupillary responses, but that this effect came at the expense of poorer memory performance on those same trials. In the current study, we observed that older adults showed the same associated costs to memory (and the additional cost of a less efficient late frontal response) but that older adults did not show the same benefits in the recovery of the N400 response in the same way that younger adults did. In other words, in noise, transient increases in effort-related arousal (measured via pupillometry) came with the same costs that younger adults experienced but there was no evidence of older adults receiving the same benefits.

In contrast to the results in noise, we observed a cost–benefit trade-off between real-time word processing and subsequent memory that was mediated by the pupillary response that emerged in quiet. In quiet, older adults were better able to recognize subsequent sentences for trials in which they had a larger pupillary response when listening to the speech. However, this increase in memory for a whole sentence came at the apparent cost to the online processing of individual words. Our results showed that for trials with larger pupil sizes, there was an increase in the onset latency of both the N400 response and the late frontal response, as well as an apparent decrease in the spatial and temporal extent of the late frontal response. Thus, in quiet, there appeared to be a pupil-mediated trade-off between subsequent recognition memory performance and the efficiency of the ERP responses, in which the listener appeared to prioritize neural resources toward sentence-level information encoding over single-word processing.

These results suggest that even at low levels of task difficulty (i.e., in the absence of acoustic challenge), older listeners had to be strategic about how to allocate cognitive and neural resources. In other words, this observed trade-off between sentence memory and the efficiency of online word-level processing indicates that the participants in the current study had some limitation on the availability of neural resources and had to be judicious in how to allocate those resources. In contrast, in previously unpublished results from the Silcox and Payne (2021) study, younger adults showed no significant relationships between pupil size and any other outcome in quiet (these results can be found here: https://osf.io/kyd6r/files

/osfstorage), suggesting that, at low levels of task difficulty, young adults do not need to be frugal in their allocation of neural resources to either online processing or memory to achieve high levels of comprehension. This qualitative difference between older and younger adults is consistent with decades of functional neuroimaging research, which has consistently found that older adults show patterns of differential recruitment of neural resources (e.g., overactivation of the frontoparietal executive function networks) at lower levels of task difficulty compared with younger adults performing the same tasks (Koen & Rugg, 2019; Sala-Llonch, Bartrés-Faz, & Junqué, 2015; Reuter-Lorenz & Park, 2014; Spaniol & Grady, 2012; Schneider-Garces et al., 2010; Park & Reuter-Lorenz, 2009; Reuter-Lorenz & Cappell, 2008; Mattay et al., 2006; Cabeza, 2002, 2004). For example, Wingfield and Grossman (2006) found that older adults showed greater frontal recruitment for linguistically simpler sentences compared with younger adults. This type of overactivation of frontal and parietal regions for relatively simple tasks is thought to reflect older adults' need to recruit executive resources to compensate for an overall reduction in neural processing efficiency. In a recently proposed model of pupillometry, Strauch and colleagues (2022) suggest that changes in pupil size not only reflect changes in the LC-NE neuromodulatory arousal network but pupil size changes may also be indirectly modulated by higher-level cortical regions that are associated with executive function, such as the frontal eye fields and the anterior cingulate cortex. Within this context, it is possible that trial-to-trial variation in pupil size that was associated with trade-offs between memory and language-related ERP responses may reflect a prioritization of neural resources driven by higher-level cortical areas responsible for executive function (which may also modulate LC-NE activity) in agreement with neuroimaging models, which suggest that older adults overactivate executive networks to compensate for age-related declines in processing efficiency at lower levels of task difficulty.

Considering the trade-offs observed in the current study that appeared to be mediated via the pupillary response, there are several implications for future research. First, the current study was specifically designed to investigate the N400 ERP component (although an exploratory analysis found evidence for a late frontal response that was also pupil-mediated), which represents only one aspect of the perceptual and cognitive processing of speech. In fact, most prior work using the ERP technique to investigate the impacts of acoustic challenge have likewise mostly focused on the N400 component (e.g., Strauß, Kotz, & Obleser, 2013; Daltrozzo, Wioland, & Kotchoubey, 2012; Obleser & Kotz, 2011; Aydelott et al., 2006; Connolly, Phillips, Stewart, & Brake, 1992). Importantly however, the processing represented by the N400 may be relatively effortless (Federmeier, 2022; Kutas & Federmeier, 2011) compared with other stages of integrative processing that take place later than the N400 and that may be more sensitive to the conscious allocation of effort (Aurnhammer, Crocker, & Brouwer, 2023; Aurnhammer, Delogu, Brouwer, & Crocker, 2023; Payne, Stites, & Federmeier, 2019; Batterink & Neville, 2013). Thus, co-registering the pupillary response with other ERP components, particularly components following the N400, may reveal more striking trade-offs in processing. Second, in speech processing research, pupillometry has primarily been used to study the impact of acoustic challenge on listening effort (for a review, see Zekveld et al., 2018). However, observing a trade-off between online speech processing and memory in a condition that had no apparent acoustic challenge in the current study highlights how this approach could be used to examine effortfulness effects beyond traditional acoustically challenging scenarios (e.g., speech-in-noise perception) to a broader range of listening environments. Finally, that we observed similar encoding-memory trade-offs without any background noise in older adults but not in younger adults also highlights the utility of adopting this pupillometry co-registration approach in investigating individual differences in such trade-offs across other groups, especially those that may be differentially impacted by limitations in cognitive and neural resources.

In conclusion, we found clear evidence that older adults engaged in context-based predictive processing that was observed in a clear N400 response, frontal effects, and multiple prediction-related benefits to memory. These effects were clearly negatively impacted by background noise as well as by trial-to-trial, pupil-mediated changes in arousal, which likely reflect attempts to overcome acoustic challenge. However, even in quiet, when task difficulty is low, trial-to-trial variation in arousal impacts real-time context processing (measured via language-related ERPs) and subsequent memory. This highlights how important it is that listening effort be conceptualized to reflect not just the impact of external task demands (e.g., changes in SNR) but should encompass internal, within-listener variation as well (Pichora-Fuller et al., 2016). The results from this study also demonstrate that using multimodal techniques (i.e., neural, pupil, and behavior) to understand listening effort are not just useful when studying young, normal-hearing adults (which is where most of this work has been done using the ERP and pupillometry techniques) but can also be used successfully for characterizing the impact of acoustic challenge and listening effort on speech comprehension in older adults.

Acknowledgments

We thank Ali Bergmark, Hope Caviness, Tyler Endow, Benning Lozada, Mark Rasmussen, and Nathania Wong for assistance with data collection on this project.

Corresponding author: Jack W Silcox, 380 South 1530 East, Salt Lake City, UT, 84106, or via e-mail: jack.silcox@utah.edu.

Data Availability Statement

The stimuli used for this study can be found at: https://osf.io/hcrv6/files/osfstorage. The data used in the analyses reported herein can be found at: https://osf.io/kyd6r/files/osfstorage.

Author Contributions

Jack W Silcox: Conceptualization; Data curation; Formal analysis; Methodology; Software; Visualization; Writing—Original draft. Karen Bennett: Data curation; Writing—Original draft. Allyson Copeland: Data curation; Writing—Original draft. Sarah Hargus Ferguson: Conceptualization; Writing—Review & editing. Brennan R. Payne: Conceptualization; Funding acquisition; Methodology; Project administration; Resources; Supervision; Writing—Original draft; Writing—Review & editing.

Funding Information

Brennan R. Payne, Center on Aging, University of Utah (https://dx.doi.org/10.13039/100020320), grant number: Pilot Award.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing*, *40*, 1084–1097. https://doi.org/10.1097/aud.0000000000000697, PubMed: 30747742

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*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709, PubMed: 16022602

Aurnhammer, C., Crocker, M. W., & Brouwer, H. (2023). Single-trial neurodynamics reveal N400 and P600 coupling in language comprehension. *Cognitive Neurodynamics*. https://doi.org/10.1007/s11571-023-09983-7

Aurnhammer, C., Delogu, F., Brouwer, H., & Crocker, M. W. (2023). The P600 as a continuous index of integration effort. *Psychophysiology*, *60*, e14302. https://doi.org/10.1111/psyp.14302, PubMed: 37042061

Ayasse, N. D., Lash, A., & Wingfield, A. (2017). Effort not speed characterizes comprehension of spoken sentences by older adults with mild hearing impairment. *Frontiers in Aging Neuroscience*, *8*, 329. https://doi.org/10.3389/fnagi.2016.00329, PubMed: 28119598

Ayasse, N. D., & Wingfield, A. (2018). A tipping point in listening effort: Effects of linguistic complexity and age-related hearing loss on sentence comprehension. *Trends in Hearing*, *22*, 2331216518790907. https://doi.org/10.1177/2331216518790907, PubMed: 30235973

Aydelott, J., Dick, F., & Mills, D. L. (2006). Effects of acoustic distortion and semantic context on event-related potentials to spoken words. *Psychophysiology*, *43*, 454–464. https://doi.org/10.1111/j.1469-8986.2006.00448.x, PubMed: 16965607

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001, PubMed: 24403724

Batterink, L., & Neville, H. J. (2013). The human brain processes syntax in the absence of conscious awareness. *Journal of Neuroscience*, *33*, 8528–8533. https://doi.org/10.1523/jneurosci.0618-13.2013, PubMed: 23658189

Benichov, J., Cox, L. C., Tun, P. A., & Wingfield, A. (2012). Word recognition within a linguistic context: Effects of age, hearing acuity, verbal ability, and cognitive function. *Ear and Hearing*, *33*, 250–256. https://doi.org/10.1097/aud.0b013e31822f680f, PubMed: 21918453

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Black, J. W. (1952). Accompaniments of word intelligibility. *Journal of Speech and Hearing Disorders*, *17*, 409–418. https://doi.org/10.1044/jshd.1704.409, PubMed: 13023794

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203–216. https://doi.org/10.1016/j.jml.2016.10.002

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, *41*, 1318–1352. https://doi.org/10.1111/cogs.12461, PubMed: 28000963

Brown, M., & Kuperberg, G. R. (2015). A hierarchical generative framework of language processing: Linking language perception, interpretation, and production abnormalities in schizophrenia. *Frontiers in Human Neuroscience*, *9*, 643. https://doi.org/10.3389/fnhum.2015.00643, PubMed: 26640435

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: The HAROLD model. *Psychology and Aging*, *17*, 85–100. https://doi.org/10.1037/0882-7974.17.1.85, PubMed: 11931290

Cabeza, R. (2004). Commentary: Neuroscience frontiers of cognitive aging: Approaches to cognitive neuroscience of aging. In R. Dixon, L. Backman, & L.-G. Nilsson (Eds.), *New frontiers in cognitive aging* (pp. 179–196). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198525691.003.0009

Carson, N., Leach, L., & Murphy, K. J. (2018). A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *International Journal of Geriatric Psychiatry*, *33*, 379–388. https://doi.org/10.1002/gps.4756, PubMed: 28731508

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, *43*, 1–18. https://doi.org/10.1016/0093-934x(92)90018-a, PubMed: 1643505

Crandell, H. A., Silcox, J. W., Ferguson, S. H., Lohani, M., & Payne, B. R. (2022). The effects of captioning errors, background noise, and hearing loss on memory for text-captioned speech. *Journal of Speech, Language, and Hearing Research*, *65*, 2364–2390. https://doi.org/10.1044/2022_jslhr-21-00416, PubMed: 35623337

Daltrozzo, J., Wioland, N., & Kotchoubey, B. (2012). The N400 and late positive complex (LPC) effects reflect controlled rather than automatic mechanisms of sentence processing. *Brain Sciences*, *2*, 267–297. https://doi.org/10.3390/brainsci2030267, PubMed: 24961195

de Gee, J. W., Colizoli, O., Kloosterman, N. A., Knapen, T., Nieuwenhuis, S., & Donner, T. H. (2017). Dynamic modulation of decision biases by brainstem arousal systems. *eLife*, *6*, e23232. https://doi.org/10.7554/elife.23232, PubMed: 28383284

Degeest, S., Keppler, H., & Corthals, P. (2015). The effect of age on listening effort. *Journal of Speech, Language, and Hearing Research*, *58*, 1592–1600. https://doi.org/10.1044/2015_jslhr-h-14-0288, PubMed: 26161899

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016, PubMed: 24953958

Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*, 1203–1207. https://doi.org/10.1111/j.1469-8986.2011.01199.x, PubMed: 21457275

Dubno, J. R., Horwitz, A. R., & Ahlstrom, J. B. (2005). Word recognition in noise at higher-than-normal levels: Decreases in scores and increases in masking. *Journal of the Acoustical Society of America*, *118*, 914–922. https://doi.org/10.1121/1.1953107, PubMed: 16158647

Ekstrom, R. B., Dermen, D., & Harman, H. H. (1976). *Manual for kit of factor-referenced cognitive tests* (Vol. 102). Princeton, NJ: Educational Testing Service.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*, 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x, PubMed: 17521377

Federmeier, K. D. (2022). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology*, *59*, e13940. https://doi.org/10.1111/psyp.13940, PubMed: 34520568

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*, 149–161. https://doi.org/10.1016/j.bandl.2010.07.006, PubMed: 20728207

Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 1–44). Elsevier Academic Press. https://doi.org/10.1016/s0079-7421(09)51001-8

Federmeier, K. D., Van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, words, sentences: Age-related changes across levels of language processing. *Psychology and Aging*, *18*, 858. https://doi.org/10.1037/0882-7974.18.4.858, PubMed: 14692871

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewalk, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101, PubMed: 16901469

Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning & Verbal Behavior*, *18*, 1–20. https://doi.org/10.1016/s0022-5371(79)90534-6

Francis, A. L., & Love, J. (2019). Listening effort: Are we measuring cognition or affect, or both? *Wiley International Reviews: Cognitive Science*, *11*, e1514. https://doi.org/10.1002/wcs.1514, PubMed: 31381275

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*, 252–269. https://doi.org/10.3758/cabn.10.2.252, PubMed: 20498349

Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *Journal of Speech, Language, and Hearing Research*, *40*, 423–431. https://doi.org/10.1044/jslhr.4002.423, PubMed: 9130210

Goy, H., Pichora-Fuller, K., van Lieshout, P., Singh, G., & Schneider, B. (2007). Effect of within- and between-talker variability on word identification in noise by younger and older adults. *Interspeech*, *2007* https://doi.org/10.21437/interspeech.2007-202

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x, PubMed: 21895683

Guang, C., Lefkowitz, E., Dillman-Hasso, N., Brown, V., & Strand, J. (2020). Recall of speech is impaired by subsequent masking noise: A replication of Rabbitt (1968) experiment 2. *Auditory Perception & Cognition*, *3*, 158–167. https://doi.org/10.1080/25742442.2021.1896908, PubMed: 34240010

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*, 1190–1192. https://doi.org/10.1126/science.143.3611.1190, PubMed: 17833905

Howes, D., & Osgood, C. E. (1954). On the combination of associative probabilities in linguistic contexts. *American Journal of Psychology*, *67*, 241–258. https://doi.org/10.2307/1418626, PubMed: 13158640

Hsin, C. H., Chao, P. C., & Lee, C. Y. (2023). Speech comprehension in noisy environments: Evidence from the predictability effects on the N400 and LPC. *Frontiers in Psychology*, *14*, 1105346. https://doi.org/10.3389/fpsyg.2023.1105346, PubMed: 36874840

Hubbard, R. J., & Federmeier, K. D. (2021). Dividing attention influences contextual facilitation and revision during language comprehension. *Brain Research*, *1764*, 147466. https://doi.org/10.1016/j.brainres.2021.147466, PubMed: 33861998

Hubbard, R. J., & Federmeier, K. D. (2024). The impact of linguistic prediction violations on downstream recognition memory and sentence recall. *Journal of Cognitive Neuroscience*, *36*, 1–23. https://doi.org/10.1162/jocn_a_02078, PubMed: 37902591

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014, PubMed: 25708148

Jongman, S. R., & Federmeier, K. D. (2022). Age-related changes in the structure and dynamics of the semantic network. *Language, Cognition, and Neuroscience*, *37*, 805–819. https://doi.org/10.1080/23273798.2021.2019286, PubMed: 36262380

Joshi, S., & Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences*, *24*, 466–480. https://doi.org/10.1016/j.tics.2020.03.005, PubMed: 32331857

Kemmotsu, N., Girard, H. M., Kucukboyaci, N. E., McEvoy, L. K., Hagler, D. J., Dale, A. M., et al. (2012). Age-related changes in the neurophysiology of language in adults: Relationship to regional cortical thinning and white matter microstructure.

Journal of Neuroscience, *32*, 12204–12213. https://doi.org/10.1523/jneurosci.0136-12.2012, PubMed: 22933802

Kiesel, A., Miller, J., Jolicoeur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, *45*, 250–274. https://doi.org/10.1111/j.1469-8986.2007.00618.x, PubMed: 17995913

Koelewijn, T., de Kluiver, H., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2015). The pupil response reveals increased listening effort when it is difficult to focus attention. *Hearing Research*, *323*, 81–90. https://doi.org/10.1016/j.heares.2015.02.004, PubMed: 25732724

Koelewijn, T., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2014). The pupil response is sensitive to divided attention during speech processing. *Hearing Research*, *312*, 114–120. https://doi.org/10.1016/j.heares.2014.03.010, PubMed: 24709275

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, *33*, 291–300. https://doi.org/10.1097/aud.0b013e3182310019, PubMed: 21921797

Koeritzer, M. A., Rogers, C. S., Van Engen, K. J., & Peelle, J. E. (2018). The impact of age, background noise, semantic ambiguity, and hearing loss on recognition memory for spoken sentences. *Journal of Speech, Language, and Hearing Research*, *61*, 740–751. https://doi.org/10.1044/2017_jslhr-h-17-0077, PubMed: 29450493

Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Jr., Cute, S. L., Humes, L. E., Dubno, J. R., et al. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*, 23–34. https://doi.org/10.1111/j.1469-8986.2012.01477.x, PubMed: 23157603

Kuchinsky, S. E., & Vaden, K. I. (2020). Aging, hearing loss, and listening effort: Imaging studies of the aging listener. In K. S. Helfer, E. L. Bartlett, A. N. Popper, & R. R. Fay (Eds.), *Aging and hearing. Springer handbook of auditory research* (Vol. 72). Cham: Springer. https://doi.org/10.1007/978-3-030-49367-7_10

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, *32*, 12–35. https://doi.org/10.1162/jocn_a_01465, PubMed: 31479347

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*, 661–679. https://doi.org/10.1017/s0140525x12003196, PubMed: 24304775

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*, 533–572. https://doi.org/10.1080/01690969308407587

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123, PubMed: 20809790

Kyong, J. S., Kwak, C., Han, W., Suh, M. W., & Kim, J. (2020). Effect of speech degradation and listening effort in reverberating and noisy environments given N400 responses. *Journal of Audiology & Otology*, *24*, 119–126. https://doi.org/10.7874/jao.2019.00514, PubMed: 32521992

Laeng, B., & Alnaes, D. (2019). Pupillometry. In *Studies in neuroscience, psychology and behavioral economics* (pp. 449–502). Springer. https://doi.org/10.1007/978-3-030-20085-5_11

Lai, M. K., Payne, B. R., & Federmeier, K. D. (2024). Graded and ungraded expectation patterns: Prediction dynamics during active comprehension. *Psychophysiology*, *61*, e14424. https://doi.org/10.1111/psyp.14424

Lai, M. K., Rommers, J., & Federmeier, K. D. (2021). The fate of the unexpected: Consequences of misprediction assessed using ERP repetition effects. *Brain Research*, *1757*, 147290. https://doi.org/10.1016/j.brainres.2021.147290, PubMed: 33516812

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*, 920–933. https://doi.org/10.1038/nrn2532, PubMed: 19020511

Luck, S. J. (2022). Artifact correction with independent component analysis. In *LibreTexts social sciences. Applied event-related potential data analysis*. University of California, Davis. https://doi.org/10.18115/D5QG92

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*, 16. https://doi.org/10.5334/joc.18, PubMed: 31517190

Mattay, V. S., Fera, F., Tessitore, A., Hariri, A. R., Berman, K. F., Das, S., et al. (2006). Neurophysiological correlates of age-related changes in working memory capacity. *Neuroscience Letters*, *392*, 32–37. https://doi.org/10.1016/j.neulet.2005.09.025, PubMed: 16213083

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology Section A*, *58*, 22–33. https://doi.org/10.1080/02724980443000151, PubMed: 15881289

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology*, *54*, 193–203. https://doi.org/10.1111/psyp.12772, PubMed: 27731503

McGarrigle, R., Knight, S., Rakusen, L., Geller, J., & Mattys, S. (2021). Older adults show a more sustained pattern of effortful listening than young adults. *Psychology and Aging*, *36*, 504. https://doi.org/10.1037/pag0000587, PubMed: 34014746

McMahon, C. M., Boisvert, I., de Lissa, P., Granger, L., Ibrahim, R., Lo, C. Y., et al. (2016). Monitoring alpha oscillations and pupil dilation across a performance-intensity function. *Frontiers in Psychology: Auditory Cognitive Neuroscience*, *7*, 745. https://doi.org/10.3389/fpsyg.2016.00745, PubMed: 27252671

Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., et al. (2017). Objective assessment of listening effort: Coregistration of pupillometry and EEG. *Trends in Hearing*, *22*, 2331216517706396. https://doi.org/10.1177/2331216517706396, PubMed: 28752807

Miller, G. A. (1951). *Language and Communication*. McGraw-Hill. https://doi.org/10.1037/11135-000

Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, *41*, 215–225. https://doi.org/10.1159/000261728, PubMed: 6535162

Murphy, D. R., Craik, F. I. M., Li, K. Z. H., & Schneider, B. A. (2000). Comparing the effects of aging and background noise on short-term memory performance. *Psychology and Aging*, *15*, 323–334. https://doi.org/10.1037/0882-7974.15.2.323, PubMed: 10879586

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, *35*, 4140–4154. https://doi.org/10.1002/hbm.22466, PubMed: 24510607

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). *Montreal Cognitive*

Assessment (MoCA). APA PsycTests. https://doi.org/10.1037/t27279-000

Navon, D. (1984). Resources–a theoretical soup stone? *Psychological Review*, *91*, 216–234. https://doi.org/10.1037/0033-295x.91.2.216

Ng, S., Payne, B. R., Steen, A. A., Stine-Morrow, E. A. L., & Federmeier, K. D. (2017). Use of contextual information and prediction by struggling adult readers: Evidence from reading times and event-related potentials. *Scientific Studies of Reading*, *21*, 359–375. https://doi.org/10.1080/10888438.2017.1310213

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, U.S.A.*, *115*, 2600–2606. https://doi.org/10.1073/pnas.1708274114, PubMed: 29531091

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage*, *55*, 713–723. https://doi.org/10.1016/j.neuroimage.2010.12.020, PubMed: 21172443

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., & Lunner, T. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hearing Research*, *365*, 90–99. https://doi.org/10.1016/j.heares.2018.05.003, PubMed: 29779607

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., et al. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research*, *351*, 68–79. https://doi.org/10.1016/j.heares.2017.05.012, PubMed: 28622894

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4, PubMed: 5146491

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity [dataset]. *Behavior Research Methods*, *47*, 1343–1355. https://doi.org/10.3758/s13428-014-0543-2, PubMed: 25479734

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: Aging and neurocognitive scaffolding. *Annual Review of Psychology*, *60*, 173–196. https://doi.org/10.1146/annurev.psych.59.103006.093656, PubMed: 19035823

Payne, B. R., & Federmeier, K. D. (2017). Pace yourself: Intraindividual variability in context use revealed by self-paced event-related brain potentials. *Journal of Cognitive Neuroscience*, *29*, 837–854. https://doi.org/10.1162/jocn_a_01090, PubMed: 28129064

Payne, B. R., & Federmeier, K. D. (2018). Contextual constraints on lexico-semantic processing in aging: Evidence from single-world event-related brain potentials. *Brain Research*, *1687*, 117–128. https://doi.org/10.1016/j.brainres.2018.02.021, PubMed: 29462609

Payne, B. R., & Silcox, J. W. (2019). Aging, context processing, and comprehension. *Psychology of Learning and Motivation*, *71*, 215–264. https://doi.org/10.1016/bs.plm.2019.07.001

Payne, B. R., Silcox, J. W., Crandell, H., Lash, A., Ferguson, S. H., & Lohani, M. (2022). Text captioning buffers against the effects of background noise and hearing loss on memory for speech. *Ear and Hearing*, *43*, 115–127. https://doi.org/10.1097/AUD.0000000000001079, PubMed: 34260436

Payne, B. R., Stites, M. C., & Federmeier, K. D. (2019). Event-related brain potentials reveal how multiple aspects of semantic processing unfold across parafoveal and foveal vision during sentence reading. *Psychophysiology*, *56*, e13432. https://doi.org/10.1111/psyp.13432, PubMed: 31274200

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*, 204–214. https://doi.org/10.1097/aud.0000000000000494, PubMed: 28938250

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37(Suppl. 1)*, 5S–27S. https://doi.org/10.1097/aud.0000000000000312, PubMed: 27355771

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, *97*, 593–608. https://doi.org/10.1121/1.412282, PubMed: 7860836

Pichora-Fuller, M. K., Schneider, B. A., MacDonald, E., Pass, H. E., & Brown, S. (2007). Temporal jitter disrupts speech intelligibility: A simulation of auditory aging. *Hearing Research*, *223*, 114–121. https://doi.org/10.1016/j.heares.2006.10.009, PubMed: 17157462

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026, PubMed: 31103785

Piquado, T., Benichov, J. I., Brownell, H., & Wingfield, A. (2012). The hidden effect of hearing acuity on speech recall, and compensatory effects of self-paced listening. *International Journal of Audiology*, *51*, 576–583. https://doi.org/10.3109/14992027.2012.684403, PubMed: 22731919

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*, 560–569. https://doi.org/10.1111/j.1469-8986.2009.00947.x, PubMed: 20070575

Rabbitt, P. M. A. (1968). Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology*, *20*, 241–248. https://doi.org/10.1080/14640746808400158, PubMed: 5683763

Rabbit, P. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Otolaryngologica*, *476*, 167–176. https://doi.org/10.3109/00016489109127274, PubMed: 2087959

Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive aging and the compensation hypothesis. *Current Directions in Psychological Science*, *17*, 177–182. https://doi.org/10.1111/j.1467-8721.2008.00570.x

Reuter-Lorenz, P. A., & Park, D. C. (2014). How does it STAC up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychology Review*, *24*, 355–370. https://doi.org/10.1007/s11065-014-9270-9, PubMed: 25143069

Richter, M., Buhiyan, T., Bramsløw, L., Innes-Brown, H., Fielder, L., Hadley, L. V., et al. (2023). Combining multiple psychophysiological measures of listening effort: Challenges and recommendations. *Seminars in Hearing*, *44*, 95–105. https://doi.org/10.1055/s-0043-1767669, PubMed: 37122882

Rogers, C. S., & Peelle, J. E. (2022). Interactions between audition and cognition in hearing loss and aging. In L. L. Holt, J. E. Peelle, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Speech perception. Springer handbook of auditory research* (Vol. 74, pp. 227–252). Cham: Springer. https://doi.org/10.1007/978-3-030-81542-4_9

Sala-Llonch, R., Bartrés-Faz, D., & Junqué, C. (2015). Reorganization of brain networks in aging: A review of functional connectivity studies. *Frontiers in Psychology*, *6*, 663. https://doi.org/10.3389/fpsyg.2015.00663, PubMed: 26052298

Schneider-Garces, N. J., Gordon, B. A., Brumback-Peltz, C. R., Shin, E., Lee, Y., Sutton, B. P., et al. (2010). Span, CRUNCH, and beyond: Working memory capacity and the aging brain. *Journal of Cognitive Neuroscience*, *22*, 655–669. https://doi.org/10.1162/jocn.2009.21230, PubMed: 19320550

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, *24*, 232–252. https://doi.org/10.1016/0749-596x(85)90026-9

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Effect of age, presentation method, and learning on identification of noise-vocoded words. *Journal of the Acoustical Society of America*, *123*, 476–488. https://doi.org/10.1121/1.2805676, PubMed: 18177175

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., et al. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*, 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526, PubMed: 28375769

Silcox, J. W., Mickey, B., & Payne, B. R. (2023). Disruption to left inferior frontal cortex modulates semantic prediction effects in reading and subsequent memory: Evidence from simultaneous TMS-EEG. *Psychophysiology*, *60*, e14312. https://doi.org/10.31219/osf.io/rphs9, PubMed: 37203307

Silcox, J. W., & Payne, B. R. (2021). The costs (and benefits) of effortful listening on context processing: A simultaneous electrophysiology, pupillometry, and behavioral study. *Cortex*, *142*, 296–316. https://doi.org/10.1016/j.cortex.2021.06.007, PubMed: 34332197

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews*: *Cognitive Science*, *5*, 679–692. https://doi.org/10.1002/wcs.1323, PubMed: 26308873

Spaniol, J., & Grady, C. (2012). Aging and the neural correlates of source memory: Over-recruitment and functional reorganization. *Neurobiology of Aging*, *33*, 425.e3–425.e18. https://doi.org/10.1016/j.neurobiolaging.2010.10.005, PubMed: 21111514

Staub, A. (2015). The effect of lexical predictability on eye movements and reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, *9*, 311–327. https://doi.org/10.1111/lnc3.12151

Stites, M. C., Payne, B. R., & Federmeier, K. D. (2017). Getting ahead of yourself: Parafoveal word expectancy modulates the N400 during sentence reading. *Cognitive, Affective, & Behavioral Neuroscience*, *17*, 475–490. https://doi.org/10.3758/s13415-016-0492-6, PubMed: 28101830

Stone, K., Nicenboim, B., Vasishth, S., & Rösler, F. (2023). Understanding the effects of constraint and predictability in ERP. *Neurobiology of Language*, *4*, 221–256. https://doi.org/10.1162/nol_a_00094

Strauch, C., Wang, C., Einhäuser, W., Van der Stigchel, S., & Naber, M. (2022). Pupillometry as an integrated readout of distinct attentional networks. *Trends in Neurosciences*, *45*, 635–647. https://doi.org/10.1016/j.tins.2022.05.003, PubMed: 35662511

Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed expectancies under degraded speech: Revisiting the N400. *Journal of Cognitive Neuroscience*, *25*, 1383–1395. https://doi.org/10.1162/jocn_a_00389, PubMed: 23489145

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring reliability. *Journalism Quarterly*, *30*, 415–433. https://doi.org/10.1177/107769905303000401

Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, *38*, 816–827. https://doi.org/10.1111/1469-8986.3850816, PubMed: 11577905

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin and Review*, *25*, 2005–2015. https://doi.org/10.3758/s13423-018-1432-y, PubMed: 29435963

Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, *41*, 167–174. https://doi.org/10.1111/j.1469-8986.2003.00148.x, PubMed: 15032982

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, *97*, 279–293. https://doi.org/10.1016/j.bandl.2005.11.003, PubMed: 16343606

Von Der Heide, R. J., Skipper, L. M., Klobusicky, E., & Olson, I. R. (2013). Dissecting the uncinate fasciculus: Disorders, controversies and a hypothesis. *Brain*, *136*, 1692–1707. https://doi.org/10.1093/brain/awt094, PubMed: 23649697

Waldron-Perrine, B., & Axelrod, B. N. (2012). Determining an appropriate cutting score for indication of impairment on the Montreal Cognitive Assessment. *International Journal of Geriatric Psychiatry*, *27*, 1189–1194. https://doi.org/10.1002/gps.3768, PubMed: 22228412

Wambacq, I., Ghanim, I., Greenfield, S., Koehnke, J., Besing, J., Chauvette, C., et al. (2023). Neurophysiologic patterns of semantic processing of accented speech. *Journal of Neurolinguistics*, *65*, 101117. https://doi.org/10.1016/j.jneuroling.2022.101117

Wendt, D., Hietkamp, R. K., & Lunner, T. (2017). Impact of noise and noise reduction on processing effort: A pupillometry study. *Ear and Hearing*, *38*, 690–700. https://doi.org/10.1097/aud.0000000000000454, PubMed: 28640038

Wendt, D., Koelewijn, T., Książek, P., Kramer, S. E., & Lunner, T. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. *Hearing Research*, *369*, 67–78. https://doi.org/10.1016/j.heares.2018.05.006, PubMed: 29858121

Wingfield, A., & Grossman, M. (2006). Language and the aging brain: Patterns of neural compensation revealed by functional brain imaging. *Journal of Neurophysiology*, *96*, 2830–2839. https://doi.org/10.1152/jn.00628.2006, PubMed: 17110737

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution of listening effort revealed by pupil dilation. *Ear and Hearing*, *36*, e153. https://doi.org/10.1097/aud.0000000000000145, PubMed: 25654299

Winn, M. B., & Teece, K. H. (2021). Listening effort is not the same as speech intelligibility score. *Trends in Hearing*, *25*, 23312165211027688. https://doi.org/10.1177/23312165211027688, PubMed: 34261392

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *Neuroimage*, *62*, 356–366. https://doi.org/10.1016/j.neuroimage.2012.04.054, PubMed: 22565202

Wlotko, E. W., Lee, C. L., & Fedeermeier, K. D. (2010). Language of the aging brain: Event-related potential studies of comprehension in older adults. Language and linguistics. *Compass*, *4*, 623–638. https://doi.org/10.1111/j.1749-818x.2010.00224.x, PubMed: 20823949

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in Hearing*, *22*, 2331216518777174. https://doi.org/10.1177/2331216518777174, PubMed: 30249172

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*, 277–284. https://doi.org/10.1111/psyp.12151, PubMed: 24506437

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing*, *31*, 480–490. https://doi.org/10.1097/aud.0b013e3181d4f251, PubMed: 20588118

Zhao, S., Bury, G., Milne, A., & Chait, M. (2019). Pupillometry as an objective measure of sustained attention in young and older listeners. *Trends in Hearing*, *23*, 2331216519887815. https://doi.org/10.1177/2331216519887815, PubMed: 31775578