

# Pupil Dilation and Microsaccades Provide Complementary Insights into the Dynamics of Arousal and Instantaneous Attention during Effortful Listening

Claudia Contadini-Wright,<sup>1</sup>  Kaho Magami,<sup>1</sup> Nishchay Mehta,<sup>2,3</sup> and  Maria Chait<sup>1</sup>

<sup>1</sup>Ear Institute, University College London, London WC1X 8EE, United Kingdom, <sup>2</sup>Hearing Health Theme, University College London Hospitals Biomedical Research Centre, London W1T 7DN, United Kingdom, and <sup>3</sup>Royal National ENT and Eastman Dental Hospitals, University College London Hospitals, London WC1E 6DG, United Kingdom

Listening in noisy environments requires effort- the active engagement of attention and other cognitive abilities- as well as increased arousal. The ability to separately quantify the contribution of these components is key to understanding the dynamics of effort and how it may change across listening situations and in certain populations. We concurrently measured two types of ocular data in young participants (both sexes): pupil dilation (PD; thought to index arousal aspects of effort) and microsaccades (MS; hypothesized to reflect automatic visual exploratory sampling), while they performed a speech-in-noise task under high- (HL) and low- (LL) listening load conditions. Sentences were manipulated so that the behaviorally relevant information (keywords) appeared at the end (Experiment 1) or beginning (Experiment 2) of the sentence, resulting in different temporal demands on focused attention. In line with previous reports, PD effects were associated with increased dilation under load. We observed a sustained difference between HL and LL conditions, consistent with increased phasic and tonic arousal. Importantly we show that MS rate was also modulated by listening load. This was manifested as a reduced MS rate in HL relative to LL. Critically, in contrast to the sustained difference seen for PD, MS effects were localized in time, specifically during periods when demands on auditory attention were greatest. These results demonstrate that auditory selective attention interfaces with the mechanisms controlling MS generation, establishing MS as an informative measure, complementary to PD, with which to quantify the temporal dynamics of auditory attentional processing under effortful listening conditions.

**Key words:** arousal; attention; eye tracking; listening effort; speech perception

## Significance Statement

Listening effort, reflecting the “cognitive bandwidth” deployed to effectively process sound in adverse environments, contributes critically to listening success. Understanding listening effort and the processes involved in its allocation is a major challenge in auditory neuroscience. Here, we demonstrate that microsaccade rate can be used to index a specific subcomponent of listening effort, the allocation of instantaneous auditory attention, that is distinct from the modulation of arousal indexed by pupil dilation (currently the dominant measure of listening effort). These results reveal the push-pull process through which auditory attention interfaces with the (visual) attention network that controls microsaccades, establishing microsaccades as a powerful tool for measuring auditory attention and its deficits.

Received Feb. 8, 2023; revised Apr. 18, 2023; accepted Apr. 26, 2023.

Author contributions: C.C.-W. and M.C. designed research; C.C.-W. performed research; K.M. and M.C. contributed unpublished reagents/analytic tools; C.C.-W. and M.C. analyzed data; C.C.-W. and M.C. wrote the first draft of the paper; C.C.-W., K.M., N.M., and M.C. edited the paper; C.C.-W. and M.C. wrote the paper.

C.C.-W. and N.M. are supported by the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre Deafness and Hearing Problems Theme.

The authors declare no competing financial interests.

Correspondence should be addressed to Maria Chait at m.chait@ucl.ac.uk or Claudia Contadini-Wright at c.wright.16@ucl.ac.uk.

<https://doi.org/10.1523/JNEUROSCI.0242-23.2023>

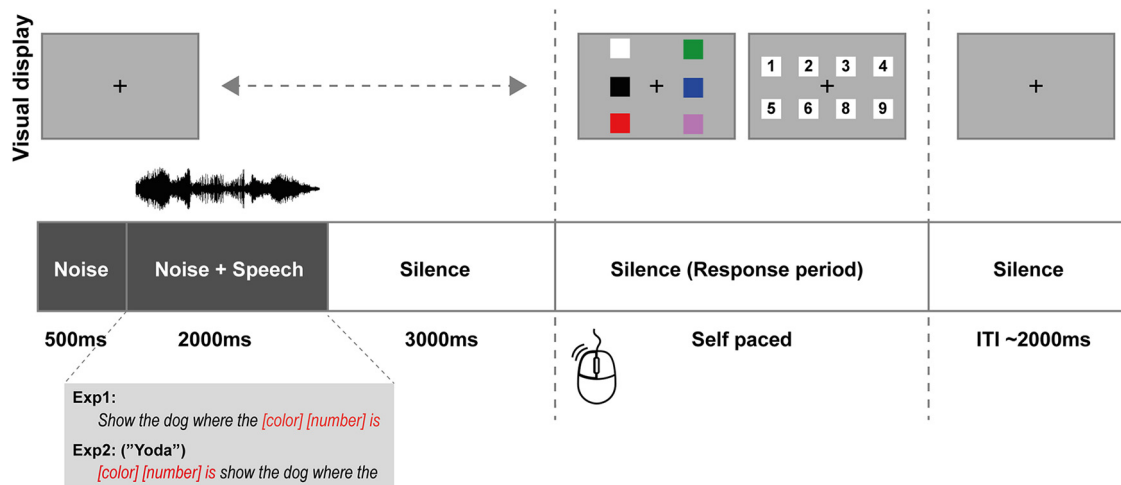
Copyright © 2023 Contadini-Wright et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

## Introduction

Listening in noisy environments draws on cognitive capacities, such as attention and memory, that support the tracking of relevant signals among interference (Peelle, 2018). The active allocation of these resources is termed “listening effort” (Pichora-Fuller et al., 2016). Better insight into listening effort and its neural underpinnings is critical for understanding the challenges faced by listeners under adverse conditions, and for managing the breakdown of these processes because of aging, hearing loss, or certain neurological conditions.

Effortful listening to speech in noise engages heightened arousal as well as increased demands on attention, working



**Figure 1.** Trial structure. In Experiment 1, sentences were in the form “show the dog where the [color] [number] is.” In Experiment 2 (“Yoda”), sentence structure was reversed such that the keywords ([color] [number]) appeared first. Trials began with 0.5 s of noise only, followed by the onset of the sentence (~2 s) and then a silent period (3 s). A response display then appeared on the screen. Participants logged their responses by selecting the correct color, then number. Visual feedback was provided. The simple sentential material and response procedure reduced demands on working memory and executive function, and emphasized the draw on attentional resources required to “fish out” the keywords from the noisy signal. Conditions of high and low listening load (blocked separately) were created by manipulating the signal (speech) to noise ratio.

memory, and linguistic processing (Peelle, 2018). Unraveling these various aspects, in particular separating arousal-related effects from those related to the allocation of cognitive resources, is a key challenge for the field (Zekveld et al., 2014; McGarrigle et al., 2021; White and Langdon, 2021; Haro et al., 2022; Ritz et al., 2022). Correlations between subjective, behavioral, and physiological measures of listening effort have not always yielded consistent results (Alhanbali et al., 2019), implying that different approaches may be capturing distinct facets of this construct.

Pupillometry is frequently used to measure the arousal component of listening effort (Ohlenforst et al., 2018; Winn et al., 2018; Zekveld et al., 2018). Non-luminance-mediated pupil dilation (PD) is linked to activity in the locus coeruleus (LC), which supplies Norepinephrine to the central nervous system and therefore controls vigilance and arousal (Wang and Munoz, 2015; Joshi et al., 2016). Stimulus-evoked PD is related to phasic activity within the LC, associated with instantaneous arousal, while baseline changes in pupil size are hypothesized to index tonic LC activity, associated with sustained alertness and engagement. Previous literature has consistently demonstrated that conditions associated with greater listening effort often lead to an increase in pupil dilation (relative to baseline; Winn et al., 2018; Zekveld et al., 2018). These effects are commonly extended in time, with peak PD, and subsequent tailing off, hypothesized to reflect a “release” of arousal at the conclusion of the perceptual process, frequently occurring after sentence offset (Winn, 2016; Winn and Moore, 2018; Winn and Teece, 2022). Conditions of high listening load have also been associated with increased baseline pupil size, proposed to reflect a balance between sustained arousal and build-up of mental fatigue (Hopstaken et al., 2015; McGarrigle et al., 2017).

Here, we hypothesize that a specific component of listening effort, the allocation of instantaneous attention, which is distinct from the state of arousal reflected by PD, can be indexed by measuring another type of ocular activity, microsaccades (MS).

Microsaccades (MS) are small fixational eye movements controlled by a network encompassing the frontal eye fields (FEFs) and the superior colliculus (SC; Hafed et al., 2015; Rucci and Victor, 2015) and are hypothesized to reflect unconscious continuous exploration of the visual environment. Recent findings

suggest that this sampling is affected by the attentional state of the individual: MS incidence reduces during, and in anticipation of, task-relevant events (Widmann et al., 2014; Denison et al., 2019; Abeles et al., 2020) and under high load (Dalmaso et al., 2017; Lange et al., 2017; Yablonski et al., 2017). Together, this evidence suggests that the processes generating MS draw on shared, limited computational capacity such that MS-indexed visual exploration is reduced when attentional resources are depleted by other perceptual tasks. Despite the richness of information potentially conveyed by MS, our understanding of how auditory perceptual processes might interface with the attentional sampling mechanisms that control MS is limited.

We recorded PD and MS concurrently while participants listened to sentences in noise under conditions of high and low listening load, modulated by manipulating the signal-to-noise ratio (SNR; Fig. 1). The placement of relevant keywords was manipulated to control the timing of instantaneous attentional engagement. We expected both PD and MS to be modulated by load, but exhibiting a different timing profile: temporally extended effects of PD-linked arousal, but transient and time-specific effects of MS incidence, precisely during periods (when keywords are presented) where the demands on auditory attention are highest.

## Materials and Methods

### Participants

All experimental procedures were approved by the research ethics committee of University College London, and written informed consent was obtained from each participant. All participants were native or long-term fluent English speakers. All reported normal hearing with no history of otological or neurological disorders. Participants reported normal or corrected-to-normal vision, with Sphere (SPH) prescriptions no greater than 3.5. A total of 70 participants were recruited and underwent testing in both experimental designs. Two different sets of participants were recruited for the two experiments.

### Experiment 1

A total of 35 paid participants between the ages of 18 and 35 took part (30 female, mean age = 23.49, SD = 4.09). Four participants were eventually excluded from the dataset, two because of substantial amounts of missing pupil data (blinks and/or gazes away from fixation; see below,

Pupillometry preprocessing and analysis) and two because of poor behavioral performance (an arbitrary threshold of <20% correct trials), resulting in a final set of 31 participants (27 female, mean age = 23.45, SD = 3.87).

### Experiment 2

A total of 35 paid participants between the ages of 18–35 took part (28 female, mean age = 25.26, SD = 5.24). Four participants were eventually excluded from the dataset, three because of substantial amounts of missing pupil data, and one because of poor behavioral performance, resulting in a final set of 31 participants (25 female, mean age = 25.45, SD = 4.97).

### Design and materials

The experimental session lasted ~1.5 h and was comprised of three stages:

1. **Threshold estimation:** A speech-in-noise reception threshold was first obtained from each participant, using the CRM task (see below, Threshold estimation). We used an adaptive procedure to determine the 50% correct threshold.
2. **Pupil screening procedures:** Prior to the main experimental session, we performed a series of brief basic measures of pupil reactivity (light reflex, dark reflex, etc.), commonly used to assess pupil function. These included measuring pupil responses to a slow, gradual change in screen brightness; to a sudden flashing white screen; to a sudden flashing black screen; and to a sudden presentation of a brief auditory stimulus (harmonic tone). These measurements were used to confirm normal pupil responsivity (Loewenfeld, 1993; Bitsios et al., 1996; Wang et al., 2018) and to identify outlying participants (none here).
3. **Main experiment:** In the main experiment, participants performed two blocks of the CRM task while their ocular data were being recorded. In one of the blocks (“High load”) the signal-to-noise ratio (SNR) was set to the threshold obtained in (1), simulating a difficult listening environment. In the second block (“Low load”), the SNR was set to the threshold obtained in (1) plus 10 dB to create a much easier listening environment (as in McGarrigle et al., 2021). The order of the two blocks was counterbalanced across participants.

All experimental tasks were implemented in MATLAB and presented via Psychophysics Toolbox version 3 (PTB-3).

### Threshold estimation

Auditory stimuli were sentences introduced by Messaoud-Galusi et al. (2011), which are a modified version of the coordinate response measure (CRM) corpus described by Bolia et al. (2000). Sentences in Experiment 1 (including threshold estimation) were in the form “Show the dog where the [color] [number] is.” Sentences in Experiment 2 (including threshold estimation) were in the form “[color] [number] is show the dog where the.” The colors that could appear within a target sentence were black, red, white, blue, green and pink. The numbers could be any digit from 1 to 9 with the exception of 7, as its bisyllabic phonetic structure makes it easier to identify. Consequently, there were a total of 48 possible combinations of color and number. Sentence duration ranged between 1.9 and 2.4 s, with the majority having a duration of 2.1 s. Sentences were embedded in Gaussian noise. The overall loudness of the noise+speech mixture was fixed at ~70 dB SPL. The SNR between the noise and speech was initially set to 20 dB, and was adjusted using a one-up-one-down adaptive procedure, tracking the 50% correct threshold. Initial steps were of 12 dB SNR, and decreased steadily following each reversal (8 dB, then 5 dB) up to a minimum step size of 2 dB. The test ended after seven reversals or after a total of 25 trials and took ~2 min to complete. The speech reception threshold was calculated as the mean SNR of the final four reversals. Participants completed 3 runs in total (the first was used as a practice). The threshold obtained from the final run was used for the “High load” condition in the main experiment. The threshold plus 10 dB was used for the “Low load” condition in the main experiment.

### Main task

In the main experiment [High load (HL) and Low load (LL) blocks; 15 min total], the same stimuli were used as for threshold estimation, but the SNR was fixed as described above. Each block contained 30 trials. Participants fixated on a black cross presented at the center of the screen (gray background). The structure of each trial is schematized in Figure 1. Trials began with 0.5 s of noise, followed by the onset of the sentence in noise (~2 s long) and then a silent period (3 s). A response display then appeared on the screen and participants logged their responses to the task by selecting the correct color first, then the number, using a mouse. Visual feedback was provided. At the end of each trial, participants were instructed to re-fixate on the cross in anticipation of the next stimulus.

### Procedure

Participants sat with their head fixed on a chinrest in front of a monitor (24-inch BENQ XL2420T with a resolution of 1920 × 1080 pixels and a refresh rate of 60 Hz) in a dimly lit and acoustically shielded room (IAC triple walled sound-attenuating booth). They were instructed to continuously fixate on a black cross presented at the center of the screen against a gray background. An infrared eye-tracking camera (Eyelink 1000 Desktop Mount, SR Research Ltd.) placed below the monitor at a horizontal distance of 62 cm from the participant was used to record pupil data. Auditory stimuli were delivered diotically through a Roland Tri-capture 24-bit 96 kHz soundcard connected to a pair of loudspeakers (Inspire T10 Multimedia Speakers, Creative Labs Inc) positioned to the left and right of the eye tracking camera. The loudness of the auditory stimuli was adjusted to a comfortable listening level for each participant. The standard five-point calibration procedure for the Eyelink system was conducted before each experimental block and participants were instructed to avoid any head movement after calibration. During the experiment, the eye-tracker continuously tracked gaze position and recorded pupil diameter, focusing binocularly at a sampling rate of 1000 Hz. Participants were instructed to blink naturally during the experiment and encouraged to rest their eyes briefly during intertrial intervals. Before each trial, the eye-tracker automatically checked that the participants’ eyes were open and fixated appropriately; trials would not start unless this was confirmed.

### Pupillometry preprocessing and analysis

Only data from the left eye was analyzed. Intervals when the participant gazed away from fixation (outside of a radius of 100 pixels around the center of the fixation cross) or when full or partial eye closure was detected (e.g., during blinks) were automatically treated as missing data. Participants with excessive missing data (>50%) were excluded from further analysis. This applied to two of the participants in Experiment 1 and three participants in Experiment 2.

### PD to speech

The pupil data from each trial were epoched from -2 to 5.5 s from sound (noise) onset. Epochs with >50% missing data were discarded from the analysis. On average, less than one trial was discarded per subject in each of the experimental blocks. Missing data in the remaining trials were recovered using linear interpolation. Trials where 10% or more of the data were identified as outlying (>3 SD from the condition mean) were removed from analysis. On average, less than one trial was discarded per subject in each of the experimental blocks. Data were then z-scored (across all trials, collapsed across the High and Low load conditions) for each participant, and time-domain averaged across all epochs of each condition to produce a single time series per condition. Both baseline-corrected (from 0.2 to 0 s preonset) and non-baseline-corrected data are reported.

### Microsaccade preprocessing and analysis

Microsaccade (MS) detection was based on an approach proposed by Engbert and Kliegl (2003). MS were extracted from the continuous eye-movement data based on the following criteria: (1) a velocity threshold of  $\lambda = 6$  times the median-based standard deviation for each subject; (2) above-threshold velocity lasting between 5 and 100 ms; (3) the events are detected in both eyes with onset disparity <10 ms; and (4) the interval

between successive microsaccades is longer than 50 ms. Extracted microsaccade events were represented as unit pulses (Dirac  $\delta$ ) and epoched as described for the PD analysis above. In Experiment 1, the eye-tracker settings (set to briefly interrupt the recording before the onset of each trial) resulted in a short period (several samples) of lost data 0.3–0 s before stimulus onset. To address the consistent artefactual absence of MS during that interval, these data were replaced by 300 ms of data between 0.6 and 0.3 s pre-onset selected from a random trial of the same participant. Eye tracker settings were modified for Experiment 2 to eliminate this issue.

In each condition, for each participant, the event time series were summed across trials and normalized by the number of trials and the sampling rate (to obtain a measure of number of events per second). Then, a causal smoothing kernel  $\omega(\tau) = \alpha^2 \times \tau \times e^{-\alpha\tau}$  was applied with a decay parameter of  $\alpha = \frac{1}{50}$  ms (Dayan and Abbott, 2001; Rolfs et al., 2008; Widmann et al., 2014), paralleling a similar technique for computing neural firing rates from neuronal spike trains (Dayan and Abbott, 2001; see also Rolfs et al., 2008). To account for the time delay caused by the smoothing kernel, the time axis was shifted by the peak of the kernel window.

### Statistical analysis

To identify time intervals in which a given pair of conditions exhibited PD/MS differences, a nonparametric bootstrap-based statistical analysis was used (Efron and Tibshirani, 1993). The difference time series between the conditions was computed for each participant and these time series were subjected to bootstrap re-sampling (1000 iterations; with replacement). At each time point, differences were deemed significant if the proportion of bootstrap iterations that fell above or below zero was  $>99\%$  (i.e.,  $p < 0.01$ ). The analysis was conducted on the full epoch as plotted and all significant time points are shown.

### Data availability

The data reported in this manuscript alongside related information are available at <https://doi.org/10.5522/04/22650472>.

## Results

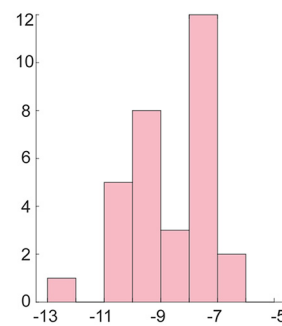
### Behavioral performance

Figure 2 shows the behavioral performance on the CRM task in Experiments 1 and 2. Speech material in Experiment 1 consisted of “standard” CRM sentences (“Show the dog where the [color] [number] is”). All sentences were initially identical (“Show the dog where the...”), allowing the listener to slowly increase arousal and attention as they prepared to identify the keywords, which always occurred at the end. In contrast, Experiment 2 used the same stimuli, but with the keywords moved to sentence onset (“[color] [number] is show the dog where the”). We refer to this condition as “Yoda” because it is similar to the speech pattern of the iconic Star Wars character (LaFrance, 2015). This material required rapid focusing of attention immediately at sentence onset, but allowed for potential disengagement of attention and arousal partway through the sentence as the remaining speech was not relevant for the task. Additionally, the unusual grammatical structure required adjusting to. Despite these distinctions, we did not observe significant differences in behavioral performance between experiments.

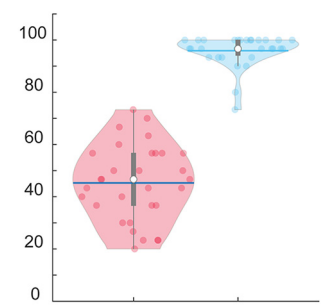
The left panels plot histograms of the 50% correct SNRs measured for each participant, these values were used for the HL condition in the main experiment. The right panels plot the performance in the main experiment (scored in %). In the HL condition, participants achieved an average of 46.34% (Experiment 1) and 46.67% (Experiment 2) correct responses, which rose to a mean of 96.02% (Experiment 1) and 96.67% (Experiment 2) in the LL condition. This confirms that the load manipulation was successful. A Mann–Whitney  $U$  test confirmed no difference

### Exp 1

#### A 50% correct thresholds

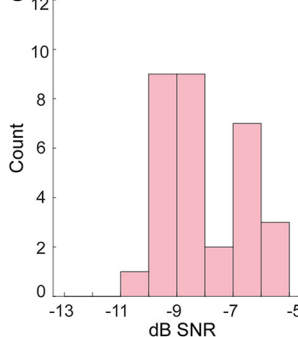


#### B Main performance

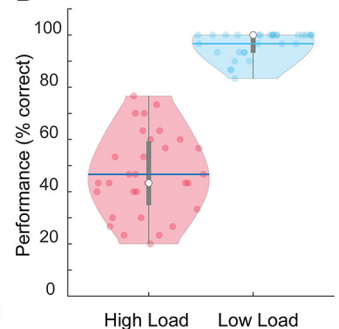


### Exp 2 ('Yoda')

#### C



#### D



**Figure 2.** Speech-in-noise performance in Experiments 1 and 2. **A, C**, Histogram of 50% correct SNRs; lower values indicated higher sensitivity to speech in noise. **B, D**, Main task performance in the High and Low Load blocks. The horizontal blue lines represent the mean, while the white dots represent the median.

between SNRs across experiments ( $U = 598$ ,  $p = 0.097$ ). Similarly, an analysis of the main performance data confirmed no main effect of experiment ( $F_{(1,60)} = 0.225$ ,  $p = 0.64$ ), only a main effect of load ( $F_{(1,60)} = 787.34$ ,  $p < 0.001$ ).

### Experiment 1: perceptual load modulates pupil size

Focusing on the load manipulation in Experiment 1, Figure 3 plots the mean PD to the target sentences in noise;  $z$  scored (across HL and LL conditions for each participant) and averaged across participants.

To quantify the sentence evoked pupil dilation, we baseline corrected the responses (0.2 s prenoise onset; Fig. 3, left). The data demonstrate a clear increase in pupil size elicited by the target sentence. In both conditions, pupil diameter rose monotonically from sentence onset, reaching its peak following sentence offset (3.09 s postonset in HL; 2.78 s postonset in LL). This is consistent with previous reports suggesting that the monotonic increase in PD reflects the gradual build-up of arousal and its subsequent release after the conclusion of perceptual processing (Winn, 2016; Winn and Moore, 2018; Winn and Teece, 2022). This increase was significantly larger in HL relative to LL, with the difference emerging from 2 s postonset and persisting to the end of the epoch. The non-baseline-corrected responses (Fig. 3, right) additionally revealed a tonic difference between HL and LL conditions that is sustained throughout the trial. This is consistent with previous observations of increased baseline pupil size under conditions of effortful listening, likely reflecting increased sustained arousal.

The effect of load on PD was observable in the majority of participants (Fig. 3, bottom) and also when analyzing only the



first five (or any random five; not shown) trials of each condition. This demonstrates that the CRM task is an effective, robust means with which to induce (and quantify) listening load. There was no correlation between task performance, quantified as the difference in performance between HL and LL conditions, and the PD difference.

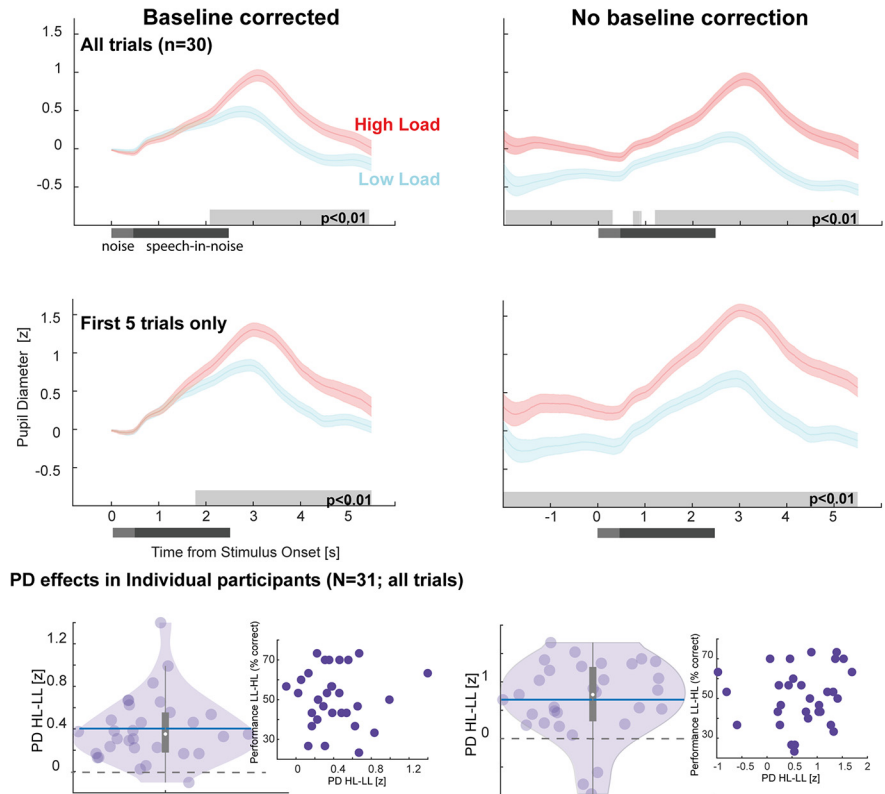
**Load-induced versus standard measures of pupil responsivity**

Figure 4 presents the load-task-related pupil responses in Experiment 1 plotted against other standard measures of pupil responsivity. These measures were obtained for each participant to confirm normal pupil reactivity and map out the responsive range, including confirming no ceiling effects. While there was individual variability in pupil size and dilation range, the plot in Figure 4 is a good representation of the average response pattern. In particular, it demonstrates that pupil responses to task-irrelevant sounds (orange trace) are tiny relative to luminance mediated responses (Pupillary light reflex and pupillary dark reflex in shades of green). Active listening (HL and LL conditions) is associated with increased tonic (baseline) and phasic (evoked by the sentence) pupil dilation. Importantly, the figure also confirms that HL pupil responses are well below ceiling as defined by the maximum pupil dilation measured for each participant.

**Experiment 1: load-induced microsaccade modulation**

Overall, the behavioral and PD data in Experiment 1 established that our task successfully manipulated auditory perceptual load, resulting in conditions that differed in listening effort. Pupil dilation data confirmed that, in line with previous reports, HL conditions were associated with sentence-evoked and sustained effects consistent with increased phasic and tonic arousal under HL. We next turn to the microsaccade rate data.

Figure 5 presents the comparison between the PD data (reproduced from Fig. 3) and microsaccade rate data (see Materials and Methods), both nonbaseline corrected. The MS data exhibit an abrupt microsaccadic inhibition (MSI) response evoked by the onset of the noise, followed by a return to baseline. MSI is commonly observed in response to abrupt sensory events and thought to reflect a rapid interruption of ongoing attentional sampling so as to prioritize the processing of a new sensory event (Rolfs et al., 2008; Roberts et al., 2019; Zhao et al., 2019). At Approximately 1.5 s postonset, both HL and LL conditions exhibit a drop in MS rate, but this effect is substantially larger for HL: a divergence between the LL and HL conditions is seen between 1.5 and 2.8 s after stimulus onset. Unlike in the PD data where a sustained difference between conditions is seen throughout the epoch, the MS rate effect is confined to a specific period, starting partway through the sentence and ending shortly after target sentence offset.



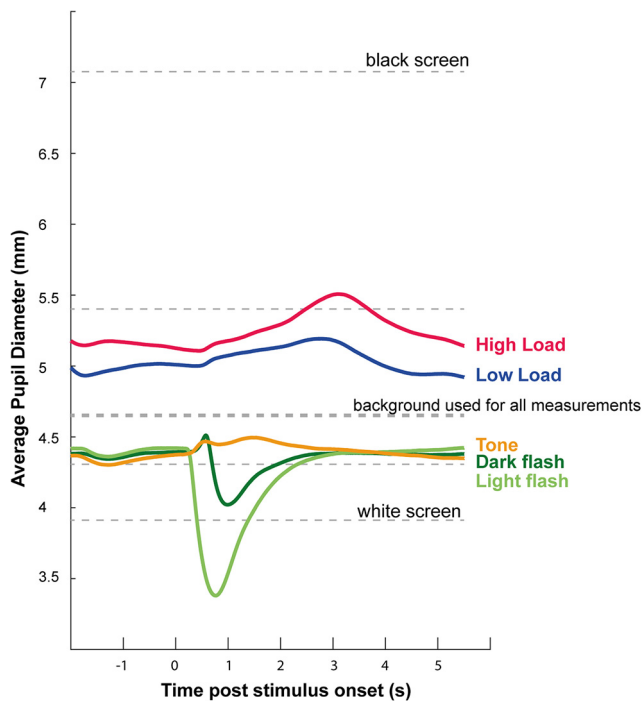
**Figure 3.** Pupil diameter is robustly modulated by listening load (Experiment 1). Top, Pupil diameter was consistently larger in the HL relative to the LL condition. This was also observable when analyzing the first five trials of each condition, confirming a robust effect. Significant differences between conditions are indicated by the gray horizontal lines. Bottom, Load effect in each participant (quantified by taking the difference between HL and LL Pupil Dilation in the interval 2–5 s postonset). Baseline-corrected data on the left; non-baseline-corrected on the right. Horizontal blue line represents the mean, while the white dot represents the median. Insets show the PD difference (horizontal axis) against performance difference between the LL and HL conditions.

In Experiment 2, we asked whether this effect is linked to the position of the behaviorally relevant elements of the sentence ([color] [number] keywords), or rather reflects the time taken for attentional effects to manifest in microsaccadic data. Unlike Experiment 1, where demands on attention (participants focusing to discern the correct color and number) were concentrated at the end of the sentence, “Yoda” sentences in Experiment 2 required focused attention at sentence onset. We hypothesized that if the lower MS rate seen in the HL condition in Experiment 1 reflects a draw on attentional resources, “Yoda” sentences would result in an earlier effect. We also hypothesized that if the timing of the PD peak (and subsequent tailing off) reflects sentence processing, its latency would similarly be shifted earlier in time in accordance with perceptual demands.

**Experiment 2: listening load in “Yoda” sentences reveals similar pupil dilation effects to those seen in Experiment 1**

“Yoda” sentences were created from the original material used in Experiment 1 by moving the keywords ([color], [number]) to the beginning of the sentence (“[color] [number] is, show the dog where the”). To succeed in the task, participants were therefore required to focus attention immediately at sentence onset, but were able to release attention/arousal resources partway through the sentence since later information was not behaviorally relevant.

Figure 6 plots the mean PD to the “Yoda” sentences in noise; z-scored (across HL and LL conditions for each participant)



**Figure 4.** Pupil responsiveness. Average (across participants) pupil responsiveness in the “pupil screening” tasks and the main (speech-in-noise) task in Experiment 1. Dashed lines, Pre-experiment measurement of pupil size at different brightness levels. Participants fixated at the center of the screen while screen brightness changed (10-s intervals; 5 brightness levels) from white to black. Mean pupil diameter in the latter 5 s of each brightness level are indicated. The screen brightness used as the default background for all other measurements is indicated by the thicker dashed line. Light green, Pupillary light reflex. Pupil responses to occasional transient (0.3 s) screen brightness changes from the default background (mid gray) to white. Average across 30 trials (7-s interflash interval). Dark green, Pupillary dark reflex. Pupil responses to occasional transient (0.3 s) screen brightness changes from the default background (mid gray) to black. Average across 30 trials (7-s interflash interval). Orange, pupil responses to occasional brief (0.3 s) harmonic tone. Average across 30 trials (7-s intertone interval). The screen brightness test was conducted first; followed by the Tone, Light, and Dark reflex tests (in random order). Blue, Main task, LL (high SNR) condition. Red, Main task, HL (low SNR) condition.

and averaged across participants. The pupil dilation responses observed in Experiment 1 were broadly replicated in Experiment 2: a greater PD was observed in the HL condition compared with the LL condition. When baseline correction was applied (Fig. 6, left) this effect was significant from  $\sim 1.5$  to  $\sim 5$  s postonset. In the nonbaseline-corrected data (Fig. 6, right) a sustained difference between conditions was also seen at baseline (from  $\sim -2$  to  $\sim -1.2$  s pre-onset), though the significant interval was somewhat shorter than that observed in Experiment 1. To elaborate on this, data from Experiment 1 and Experiment 2 were compared directly (Fig. 7C).

### Sentence structure modulates the PD response

Following previous observations that the timing of the PD peak (and subsequent tailing off) tends to occur after sentence offset, reflecting release of arousal at the conclusion of perceptual processing (Winn, 2016; Winn and Moore, 2018; Winn and Teece, 2022), we hypothesized that the latency of the peak PD in Experiment 2, relative to Experiment 1, would be shifted in accordance with the change in sentence structure. Figure 7 compares the mean PD data across Experiments 1 and 2. A comparison of the HL conditions (nonbaseline corrected; Fig. 7A) revealed a similar baseline in both experiments. This might

be interpreted to suggest a similar effortful state (we return to this point in the Discussion section). As predicted, Experiment 2 saw the latency of the PD peak shift earlier in time. The comparison between the low load conditions (Fig. 7B) yielded a similar overall pattern to that observed for HL.

We also observed differences between Experiments 1 and 2 in the peak amplitude of the PD in the HL condition. An independent samples *t* test revealed a significant difference in peak amplitude in the HL condition ( $t_{(60)} = 2.29, p = 0.025$ ) but not in the LL condition ( $t_{(60)} = -0.48, p = 0.63$ ). The pattern of the PD overall suggests a gradual increase in arousal from sentence onset in both experiments with an earlier release (and a lower peak) in Experiment 2.

Figure 7C quantifies the difference between HL and LL in the two experiments by plotting the (baseline corrected) difference waveforms. In Experiment 1, a divergence between HL and LL is seen from  $\sim 2$  s post-onset. In Experiment 2, the two conditions diverge 0.5 s earlier (at  $\sim 1.5$  s postonset), again consistent with the earlier engagement of arousal in Experiment 2.

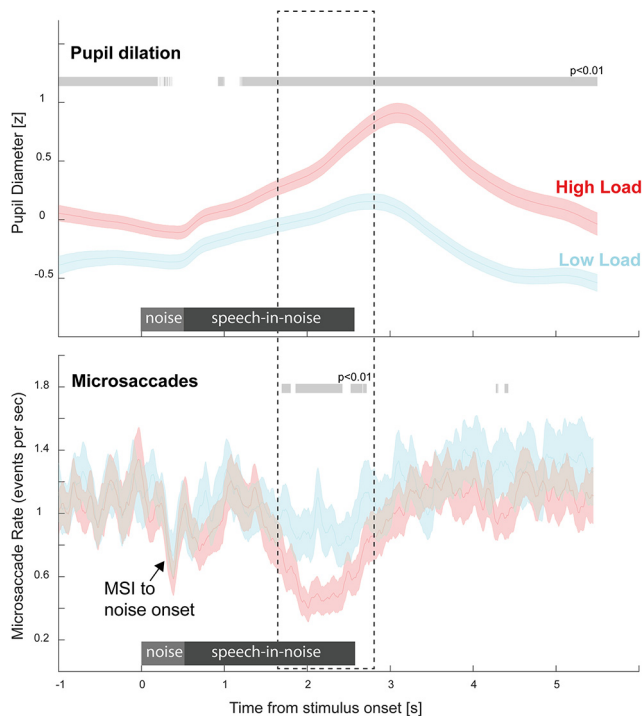
### Experiment 2: load-induced microsaccade modulation

Figure 8 presents the comparison between the PD data (reproduced from Fig. 6) and microsaccade rate data, both nonbaseline corrected. The MS data exhibit an abrupt microsaccadic inhibition (MSI) response evoked by the onset of the noise. Unlike Experiment 1, this response is not followed by a return to baseline. Microsaccade rate remains low until partway through the sentence in both HL and LL conditions. Thereafter, a return to baseline is observed. This is consistent with the fact that in Experiment 2 attentional resources are required during the initial portion of the sentence. In the LL condition, this inhibition returns to baseline faster than in the HL condition, resulting in a significant difference around  $\sim 1.2$ – $1.5$  s post-onset. Thus, while a reduction in MS rate was present in both conditions, it lasted longer in the HL condition.

Furthermore, there appears to be an additional (weak) effect during the baseline (pre-sentence onset) period where the LL condition exhibits a reduction in MS rate from  $\sim 500$  ms before sound onset. This is an incidental finding, and as such must be clarified in follow-up investigations. The preonset effects might reflect the anticipatory preallocation of attentional resources. That the effect is only observed in the LL condition is interesting and might imply a depletion of relevant resources in HL that precluded participants from preparing to attend in the same way. We return to this in the Discussion section.

### Microsaccade rate modulation reflects localized demand on attention

To directly compare the timing of MS rate modulation, Figure 9 presents data from Experiments 1 and 2 separately for the HL and LL conditions. MS rate was significantly lower in Experiment 2 during the initial portion of the sentence and vice versa during the latter portion of the sentence – consistent with where the keywords were embedded. This effect is seen in both the HL and LL conditions, but greater (in terms of deflection from baseline) in the HL condition, reflecting the higher demand on attentional resources during low SNR. Critically, in each condition the period of MS modulation (where MS rate was different from baseline) was confined to the period in the sentence where the keywords were present. See Extended Data Figure 9-1 for an analysis of associated



**Figure 5.** MS incidence is modulated by listening load (Experiment 1). Concurrently recorded PD and MS data. Data are not baseline corrected. Significant differences between conditions are indicated by the gray horizontal lines.

blink rates, confirming that MS modulation is not a consequence of increased blinking during the relevant time periods.

## Discussion

Successful listening in busy environments requires the deployment of arousal, attention, and other cognitive functions, usually collated under the umbrella-term “listening effort” (Peelle, 2018; Pichora-Fuller et al., 2016). Quantifying the contribution of each factor to effortful listening is critical for understanding the challenges listeners face in adverse conditions, characterizing individual difficulties (Winn and Teece, 2022) and guiding rehabilitation. We measured pupil dilation (PD) and microsaccade (MS) rate while listeners performed a speech-in-noise detection task simulating high- and low-listening load conditions (HL and LL). PD is a prevalent measure of listening effort (Zekveld et al., 2018; Winn et al., 2018), believed to index modulation of arousal. MS are hypothesized to reflect a process of unconscious visual exploratory sampling, and are therefore a potentially useful signal for understanding how the auditory system interfaces with the brain’s attention network. We hypothesized that measuring MS alongside PD would enable us to precisely identify and measure the deployment of focused attention during speech processing.

Consistent with a large body of work on PD as an index of listening effort (Ohlenforst et al., 2018; Winn et al., 2018; Zekveld et al., 2018), we observed phasic (sentence evoked) and tonic (sustained) differences in pupil diameter between HL and LL conditions. Unlike the temporally extended PD responses, MS effects, characterized by a reduced MS rate, were only present when demands on focused attention were highest. These results demonstrate that auditory selective attention interfaces with the mechanisms controlling MS generation, establishing

MS as an informative measure with which to quantify the temporal dynamics of auditory attentional processing during effortful listening.

### Pupil dilation data reveal tonic and phasic modulation of arousal under load

In line with the link between PD and listening effort and consistent with the general association between greater arousal and larger pupil sizes (Partala and Surakka, 2003; Bradley et al., 2008), we observed robust effects of load on the pupillary response. Baseline-corrected PD data revealed a significant difference between load conditions, suggesting greater instantaneous arousal under HL. Additional extended differences in nonbaseline-corrected data are consistent with sustained alertness and engagement under HL.

The latency of the PD peak, and subsequent tailing off, shifted with the positions of the behaviorally relevant keywords (Fig. 7). This is consistent with the notion of “effort release” (Winn, 2016; Winn and Moore, 2018; Winn and Teece, 2021; Winn and Teece, 2022). Winn (2016; see also McMurray et al., 2017) demonstrated that sentential material of increasing complexity (low semantic context or vocoding) is associated with later decay of PD post-offset, indicative of increased (and presumably more temporally prolonged) demands on processing lasting beyond sentence offset. In Winn and Moore (2018), effort release happened earlier when sentences were followed by ignored stimuli compared with attended digit sequences. Similar effects were observed here with the PD peak occurring earlier for the “Yoda” stimuli (Experiment 2) relative to the original sentences (Experiment 1).

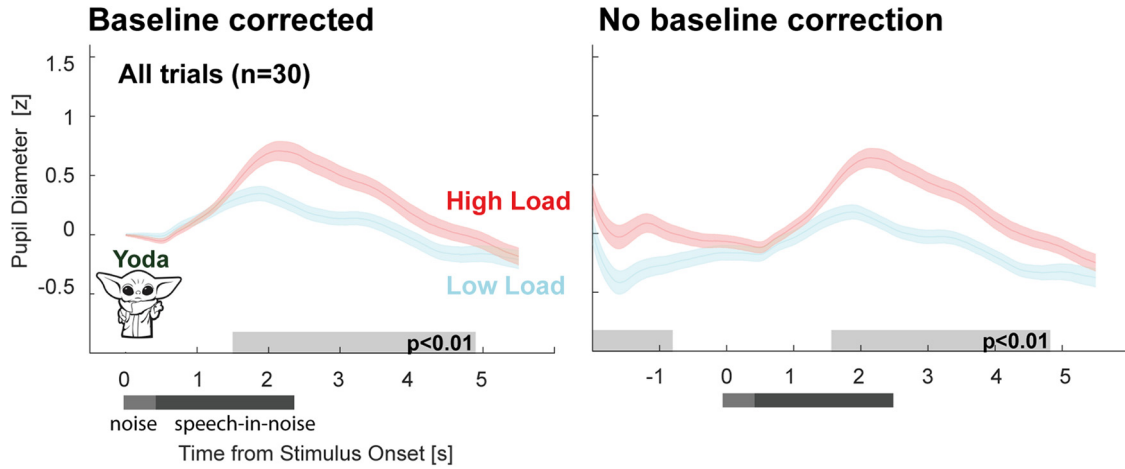
In addition to the shift in latency we also observed that the amplitude of the HL peak in Experiment 1 was larger than that in Experiment 2 (Fig. 7A). This reveals that effortful listening is associated with a steady accumulation of arousal as the sentence unfolds. Hence, earlier behavioral disengagement (e.g., “Yoda” sentences here) is associated with both an earlier and shallower peak dilation.

In summary, several processes, instantaneous arousal, attentional engagement, and sustained arousal related to the adverse listening environment all appear to contribute to observed PD modulation. Notably, PD effects are sluggish relative to the timing of behaviorally relevant information. This may reflect slow changes in arousal, or instead delays in the circuit linking modulation of arousal and pupil musculature.

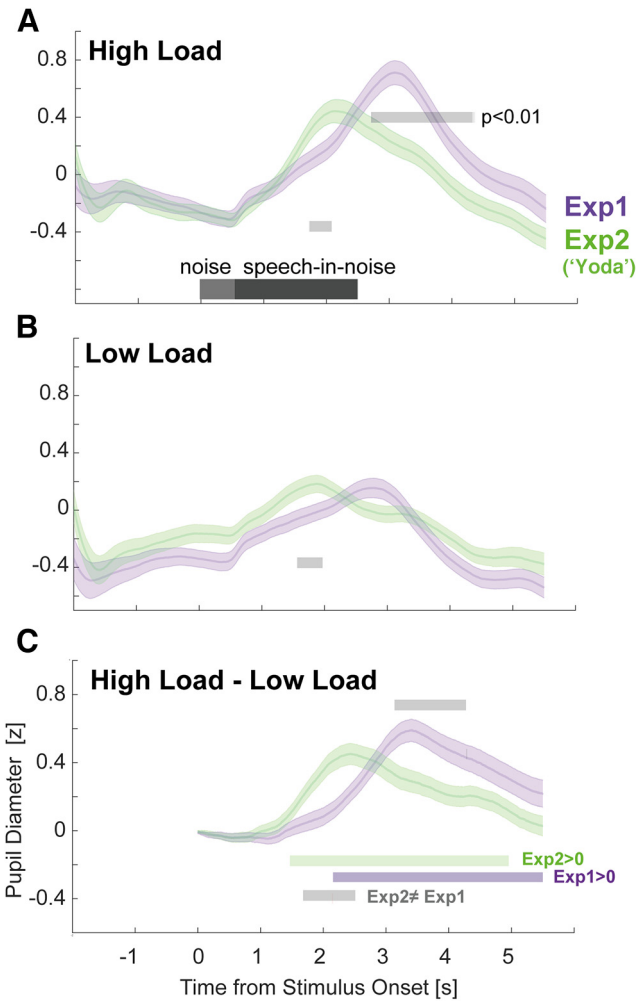
### Microsaccadic activity indexes instantaneous auditory attention

Microsaccades are small spontaneous fixational eye movements occurring at a rate between 1 and 2 Hz. Accumulating evidence suggests that MS reflect a process of unconscious visual exploratory sampling which modulates early visual processing and plays a critical role in visual perception (Lowet et al., 2018). Importantly, recent findings suggest that MS sampling draws on a central resource pool shared with other perceptual processes, such that MS incidence is affected by the load currently experienced by the individual. MS-indexed visual exploration has been shown to reduce in anticipation of task-relevant events (Denison et al., 2019; Abeles et al., 2020), as a function of task engagement (e.g., absorption during music listening; Lange et al., 2017), and under high load (Siegenthaler et al., 2014; Gao et al., 2015; Dalmaso et al., 2017; Yablonski et al., 2017). MS is therefore a useful signal for quantifying participants’ instantaneous attentional state.

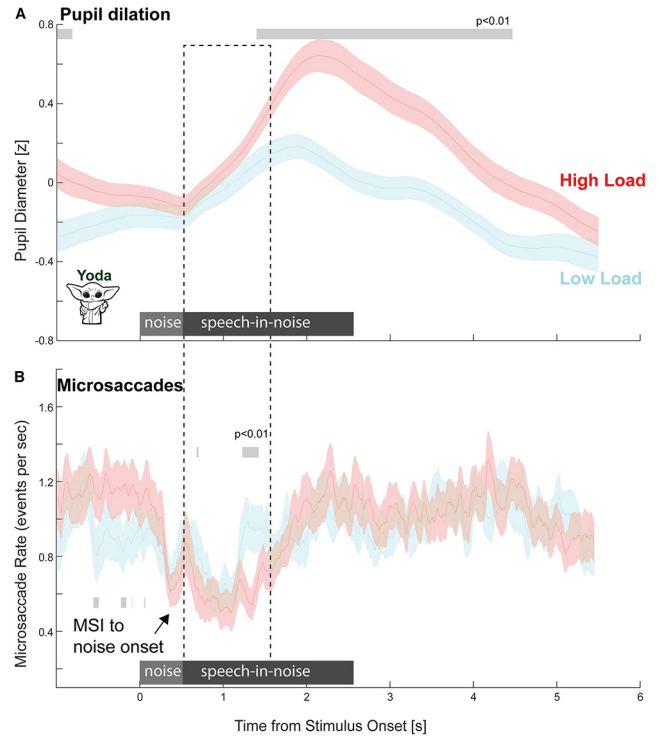




**Figure 6.** Experiment 2. Pupil diameter is modulated by listening load in ‘Yoda’ sentences. Pupil diameter was consistently larger in the HL relative to the LL condition. Significant differences between conditions are indicated by the gray horizontal lines. Baseline-corrected data on the left; nonbaseline-corrected on the right. These results replicated the findings of Experiment 1.



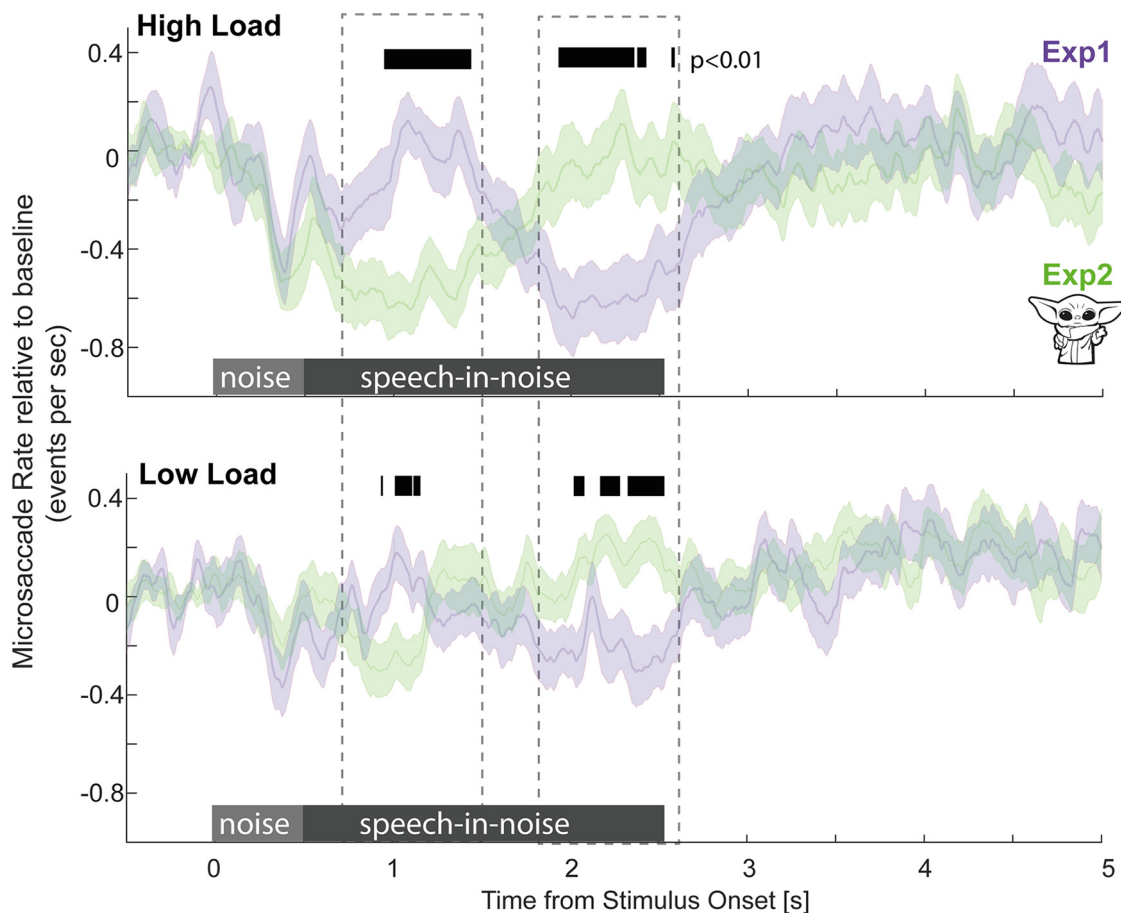
**Figure 7.** Sentence structure modulates latency of PD response. **A**, PD in the HL condition across experiments. Significant differences between experiments are indicated by the gray horizontal lines. **B**, PD in the LL condition across experiments. Significant differences between experiments are indicated by the gray horizontal lines. **C**, HL-LL Difference between experiments. Green horizontal line, Experiment 2 difference when compared with zero. Purple horizontal line, Experiment 1 difference when compared with zero.



**Figure 8.** Experiment 2: MS incidence is modulated by listening load. Concurrently recorded PD and MS data. Data are not baseline corrected. Significant differences between conditions are indicated by the gray horizontal lines.

Although research into MS dynamics has predominantly focused on visual processing, spatial attention in particular (Engbert, 2006; Krauzlis et al., 2017), previous demonstrations have linked MS-indexed sampling and auditory processing: anticipation of target sounds causes reduced sustained MS activity (Widmann et al., 2014; Abeles et al., 2020) and perceptual salience of brief sounds modulates transient evoked microsaccadic inhibition (Valsecchi and Turatto, 2009; Zhao et al., 2019). Interestingly, abrupt sounds cause more rapid MS inhibition than visual stimuli (Rolfs, 2009), suggesting fast circuitry consistent with the auditory system having privileged access into MS-indexed attention. Here, we further demonstrated





**Figure 9.** MS rate reflects instantaneous demand on attentional resources. Top, HL conditions in Experiment 1 and Experiment 2 (“Yoda”). Bottom, LL conditions in Experiment 1 and Experiment 2 (“Yoda”), all baseline corrected. Significant differences between conditions are indicated with the black horizontal bar. See Extended Data Figure 9-1 for an analysis of blink rates.

that instantaneous top-down auditory attention modulates MS dynamics.

Microsaccade activity was recorded concurrently with PD. Since no element of the task was spatialized, we pooled over MS direction and focused on incidence. Unlike the sustained differences between HL and LL conditions observed in the PD data, MS dynamics exhibited a localized effect specifically at points within the sentence containing behaviorally relevant information (keywords). This supports the hypothesis that MS rate reduction reflects an instantaneous draw on attentional resources. A similar pattern of MS dynamics was observed in HL and LL conditions (Fig. 9), but the effects relative to baseline are more distinct in HL. This reflects a greater draw on attentional resources exacerbated by the arduous listening environment. Overall, these results demonstrate that MS dynamics can track participants’ instantaneous attentional state: monitoring microsaccade activity during speech processing can reveal how listeners allocate attentional resources to the unfolding sentential material.

In Experiment 2 (Yoda sentences), a smaller difference in MS rate between HL and LL was observed, potentially because the unusual listening demands (keywords exactly at onset) depleted attentional resources in both conditions. Interestingly, while MS modulation was largely confined to keyword locations, some pre-onset effects were seen in the LL condition, potentially reflecting preparatory attention allocation to sentence onset. Curiously, this effect was not observed in HL, perhaps because load-induced fatigue depleted listeners’ ability to effectively deploy preparatory

attention. This incidental observation could be expanded on in future research to gain a fine-grained perspective on the dynamics of auditory attention.

### Unraveling arousal and attention effects

The MS and PD effects reported here reflect the coordinated operation of a network that regulates arousal and attention. However, the architecture of this network, and specifically the interrelation between PD and MS control circuits, are only vaguely elaborated.

MS are controlled by a network encompassing the superior colliculus (SC) and frontal eye field (FEF), an area that plays a key role in controlling attention and distraction (Peel et al., 2016; Lega et al., 2019; Hsu et al., 2021). The SC is considered the dominant driver of MS (Hafed, 2011; Goffart et al., 2012; Hafed and Krauzlis, 2012) and a key structure for visual attention (Herman and Krauzlis, 2017). It receives sensory, cognitive, and arousal inputs from cortical and subcortical sources (including from the locus coeruleus (LC); see below) and projects to brainstem premotor circuits to direct the orienting response.

PD is commonly linked to LC activity (Aston-Jones and Cohen, 2005; Murphy et al., 2014; Joshi et al., 2016; Reimer et al., 2016; Breton-Provencher and Sur, 2019) which is at the core of arousal regulation (Aston-Jones and Cohen, 2005; Berridge, 2008; Samuels and Szabadi, 2008a, b; Breton-Provencher and Sur, 2019). The LC has also been implicated in attentional control (Robbins, 1984; Aston-Jones et al., 1999) through projections to the prefrontal cortex (PFC) and SC (Wang et al., 2017; Benavidez et al., 2021; Wang

and Munoz, 2021). Indeed, LC-NE system responsiveness correlates with enhanced attentional performance (Dahl et al., 2020). Optogenetic activation of LC-NE neurons improves attention and response inhibition via projections to PFC (Bari et al., 2020), with overall attentional state mediated by tonic and phasic LC activity (Aston-Jones and Cohen, 2005; Howells et al., 2012). Therefore, the PD signal likely indexes activity correlating with both attention and arousal.

In line with the multiple links between MS and PD circuits (including via SC and PFC), there are reports of correlations between PD-indexed micro fluctuations of arousal and spontaneous, or SC microstimulation-evoked, MS activity (Wang and Munoz, 2021; Johnston et al., 2022; Wang et al., 2022). Similarly, here the modulation of MS rate occurred during the rising slope of PD (where change in pupil size was fastest). This is consistent with attention and arousal interacting to support the perceptual processing of the sentence.

Overall, our results show that while the pupil signal is affected by both attention and arousal, the time specificity of MS modulation renders them a powerful tool for pinpointing effects of attention. Future work can capitalize on MS as a critical window into the dynamics of auditory focused attention and its deficits.

## References

- Abeles D, Amit R, Tal-Perry N, Carrasco M, Yuval-Greenberg S (2020) Oculomotor inhibition precedes temporally expected auditory targets. *Nat Commun* 11:3524.
- Alhanbali S, Dawes P, Millman RE, Munro KJ (2019) Measures of listening effort are multidimensional. *Ear Hear* 40:1084–1097.
- Aston-Jones G, Cohen JD (2005) Adaptive gain and the role of the locus coeruleus–norepinephrine system in optimal performance. *J Comp Neurol* 493:99–110.
- Aston-Jones G, Rajkowski J, Cohen J (1999) Role of locus coeruleus in attention and behavioral flexibility. *Biol Psychiatry* 46:1309–1320.
- Bari A, Xu S, Pignatelli M, Takeuchi D, Feng J, Li Y, Tonegawa S (2020) Differential attentional control mechanisms by two distinct noradrenergic coeruleo-frontal cortical pathways. *Proc Natl Acad Sci USA* 117:29080–29089.
- Benavidez NL, et al. (2021) Organization of the inputs and outputs of the mouse superior colliculus. *Nat Commun* 12:4004.
- Berridge CW (2008) Noradrenergic modulation of arousal. *Brain Res Rev* 58:1–17.
- Bitsios P, Prettyman R, Szabadi E (1996) Changes in autonomic function with age: a study of pupillary kinetics in healthy young and old people. *Age Ageing* 25:432–438.
- Bolia RS, Nelson WT, Ericson MA, Simpson BD (2000) A speech corpus for multitalker communications research. *J Acoust Soc Am* 107:1065–1066.
- Bradley MM, Miccoli L, Escrig MA, Lang PJ (2008) The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45:602–607.
- Breton-Provencher V, Sur M (2019) Active control of arousal by a locus coeruleus GABAergic circuit. *Nat Neurosci* 22:218–228.
- Dahl MJ, Mather M, Sander MC, Werkle-Bergner M (2020) Noradrenergic responsiveness supports selective attention across the adult lifespan. *J Neurosci* 40:4372–4390.
- Dalmaso M, Castelli L, Scatturin P, Galfano G (2017) Working memory load modulates microsaccadic rate. *J Vis* 17:6.
- Dayan P, Abbott LF (2001) Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge: Massachusetts Institute of Technology Press.
- Denison RN, Yuval-Greenberg S, Carrasco M (2019) Directing voluntary temporal attention increases fixational stability. *J Neurosci* 39:353–363.
- Efron B, Tibshirani R (1993) An introduction to the bootstrap. New York: Chapman and Hall.
- Engbert R (2006) Microsaccades: a microcosm for research on oculomotor control, attention, and visual perception. *Prog Brain Res* 154:177–192.
- Engbert R, Kliegl R (2003) Microsaccades uncover the orientation of covert attention. *Vision Res* 43:1035–1045.
- Gao X, Yan H, Sun H (2015) Modulation of microsaccade rate by task difficulty revealed through between- and within-trial comparisons. *J Vis* 15:3.
- Goffart L, Hafed ZM, Krauzlis RJ (2012) Visual fixation as equilibrium: evidence from superior colliculus inactivation. *J Neurosci* 32:10627–10636.
- Hafed ZM (2011) Mechanisms for generating and compensating for the smallest possible saccades. *Eur J Neurosci* 33:2101–2113.
- Hafed ZM, Krauzlis RJ (2012) Similarity of superior colliculus involvement in microsaccade and saccade generation. *J Neurophysiol* 107:1904–1916.
- Hafed ZM, Chen CY, Tian X (2015) Vision, perception, and attention through the lens of microsaccades: mechanisms and implications. *Front Syst Neurosci* 9:167.
- Haro S, Rao HM, Quatieri TF, Smalt CJ (2022) EEG alpha and pupil diameter reflect endogenous auditory attention switching and listening effort. *Eur J Neurosci* 55:1262–1277.
- Herman JP, Krauzlis RJ (2017) Color-change detection activity in the primate superior colliculus. *eNeuro* 4:ENEURO.0046-17.2017.
- Hopstaken JF, van der Linden D, Bakker AB, Kompier MAJ (2015) The window of my eyes: task disengagement and mental fatigue covary with pupil dynamics. *Biol Psychol* 110:100–106.
- Howells FM, Stein DJ, Russell VA (2012) Synergistic tonic and phasic activity of the locus coeruleus norepinephrine (LC-NE) arousal system is required for optimal attentional performance. *Metab Brain Dis* 27:267–274.
- Hsu TY, Chen JT, Tseng P, Wang CA (2021) Role of the frontal eye field in human microsaccade responses: a TMS study. *Biol Psychol* 165:108202.
- Johnston R, Snyder AC, Khanna SB, Issar D, Smith MA (2022) The eyes reflect an internal cognitive state hidden in the population activity of cortical neurons. *Cereb Cortex* 32:3331–3346.
- Joshi S, Li Y, Kalwani RM, Gold JI (2016) Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron* 89:221–234.
- Krauzlis RJ, Goffart L, Hafed ZM (2017) Neuronal control of fixation and fixational eye movements. *Phil Trans R Soc B* 372:20160205.
- LaFrance A (2015) An unusual way of speaking. Yoda has. Atlantic. Available at: <https://www.theatlantic.com/entertainment/archive/2015/12/hmmmmm/420798/>. Retrieved February 1, 2023.
- Lange EB, Zweck F, Sinn P (2017) Microsaccade-rate indicates absorption by music listening. *Conscious Cogn* 55:59–78.
- Lega C, Ferrante O, Marini F, Santandrea E, Cattaneo L, Chelazzi L (2019) Probing the neural mechanisms for distractor filtering and their history-contingent modulation by means of TMS. *J Neurosci* 39:7591–7603.
- Loewenfeld IE (1993) The pupil: anatomy, physiology, and clinical applications. Ames: Iowa State University Press.
- Lowet E, Gomes B, Srinivasan K, Zhou H, Schafer RJ, Desimone R (2018) Enhanced neural processing by covert attention only during microsaccades directed toward the attended stimulus. *Neuron* 99:207–214.e3.
- McGarrigle R, Dawes P, Stewart AJ, Kuchinsky SE, Munro KJ (2017) Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology* 54:193–203.
- McGarrigle R, Rakusen L, Mattys S (2021) Effortful listening under the microscope: examining relations between pupillometric and subjective markers of effort and tiredness from listening. *Psychophysiology* 58:e13703.
- McMurray B, Farris-Trimble A, Rigler H (2017) Waiting for lexical access: cochlear implants or severely degraded input lead listeners to process speech less incrementally. *Cognition* 169:147–164.
- Messaoud-Galusi S, Hazan V, Rosen S (2011) Investigating speech perception in children with dyslexia: is there evidence of a consistent deficit in individuals? *J Speech Lang Hear Res* 54:1682–1701.
- Murphy PR, O'Connell RG, O'Sullivan M, Robertson IH, Balsters JH (2014) Pupil diameter covaries with BOLD activity in human locus coeruleus. *Hum Brain Mapp* 35:4140–4154.
- Ohlenforst B, Wendt D, Kramer SE, Naylor G, Zekveld AA, 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. *Hear Res* 365:90–99.
- Partala T, Surakka V (2003) Pupil size variation as an indication of affective processing. *Int J Hum Comput Stud* 59:185–198.
- Peel TR, Hafed ZM, Dash S, Lomber SG, Corneil BD (2016) A causal role for the cortical frontal eye fields in microsaccade deployment. *PLoS Biol* 14:e1002531.

- Peelle JE (2018) Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear* 39:204–214.
- Pichora-Fuller MK, Kramer SE, Eckert MA, Edwards B, Hornsby BWY, Humes LE, Lemke U, Lunner T, Matthen M, Mackersie CL, Naylor G, Phillips NA, Richter M, Rudner M, Sommers MS, Tremblay KL, Wingfield A (2016) Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear Hear* 37:5S–27S.
- Reimer J, McGinley MJ, Liu Y, Rodenkirch C, Wang Q, McCormick DA, Tollas AS (2016) Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat Commun* 7:13289.
- Ritz H, Wild CJ, Johnsrude IS (2022) Parametric cognitive load reveals hidden costs in the neural processing of perfectly intelligible degraded speech. *J Neurosci* 42:4619–4628.
- Robbins TW (1984) Cortical noradrenaline, attention and arousal. *Psychol Med* 14:13–21.
- Roberts MJ, Lange G, Van Der Veen T, Lowet E, De Weerd P (2019) The attentional blink is related to the microsaccade rate signature. *Cereb Cortex* 29:5190–5203.
- Rolfes M (2009) Microsaccades: small steps on a long way. *Vision Res* 49:2415–2441.
- Rolfes M, Kliegl R, Engbert R (2008) Toward a model of microsaccade generation: the case of microsaccadic inhibition. *J Vis* 8:5.
- Rucci M, Victor JD (2015) The unsteady eye: an information-processing stage, not a bug. *Trends Neurosci* 38:195–206.
- Samuels ER, Szabadi E (2008a) Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part I: principles of functional organisation. *Curr Neuropharmacol* 6:235–253.
- Samuels ER, Szabadi E (2008b) Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part II: physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. *Curr Neuropharmacol* 6:254–285.
- Siegenthaler E, Costela FM, McCamy MB, Di Stasi LL, Otero-Millan J, Sonderegger A, Groner R, Macknik S, Martinez-Conde S (2014) Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *Eur J Neurosci* 39:287–294.
- Valsecchi M, Turatto M (2009) Microsaccadic responses in a bimodal oddball task. *Psychol Res* 73:23–33.
- Wang CA, Munoz DP (2015) A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. *Curr Opin Neurobiol* 33:134–140.
- Wang CA, Blohm G, Huang J, Boehnke SE, Munoz DP (2017) Multisensory integration in orienting behavior: pupil size, microsaccades, and saccades. *Biol Psychol* 129:36–44.
- Wang CA, Tworzynski L, Huang J, Munoz DP (2018) Response anisocoria in the pupillary light and darkness reflex. *Eur J Neurosci* 48:3379–3388.
- Wang C-A, Munoz DP (2021) Differentiating global luminance, arousal and cognitive signals on pupil size and microsaccades. *Eur J Neurosci* 54:7560–7574.
- Wang CA, White B, Munoz DP (2022) Pupil-linked arousal signals in the midbrain superior colliculus. *J Cogn Neurosci* 34:1340–1354.
- White BE, Langdon C (2021) The cortical organization of listening effort: new insight from functional near-infrared spectroscopy. *Neuroimage* 240:118324.
- Widmann A, Engbert R, Schröger E (2014) Microsaccadic responses indicate fast categorization of sounds: a novel approach to study auditory cognition. *J Neurosci* 34:11152–11158.
- Winn MB (2016) Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends Hear* 20:233121651666972.
- Winn MB, Moore AN (2018) Pupillometry reveals that context benefit in speech perception can be disrupted by later-occurring sounds, especially in listeners with cochlear implants. *Trends Hear* 22:2331216518808962.
- Winn MB, Teece K (2021) The timing of deploying and withholding listening effort, in listeners with normal hearing or with cochlear implants. *J Acoust Soc Am* 150:A340.
- Winn MB, Teece KH (2022) Effortful listening despite correct responses: the cost of mental repair in sentence recognition by listeners with cochlear implants. *J Speech Lang Hear Res* 65:3966–3980.
- Winn MB, Wendt D, Koelewijn T, Kuchinsky SE (2018) Best practices and advice for using pupillometry to measure listening effort: an introduction for those who want to get started. *Trends Hear* 22:2331216518800869.
- Yablonski M, Polat U, Bonneh YS, Ben-Shachar M (2017) Microsaccades are sensitive to word structure: a novel approach to study language processing. *Sci Rep* 7:3999.
- Zekveld AA, Heslenfeld DJ, Johnsrude IS, Versfeld NJ, Kramer SE (2014) The eye as a window to the listening brain: neural correlates of pupil size as a measure of cognitive listening load. *Neuroimage* 101:76–86.
- Zekveld AA, Koelewijn T, Kramer SE (2018) The pupil dilation response to auditory stimuli: current state of knowledge. *Trends Hear* 22:2331216518777174.
- Zhao S, Yum NW, Benjamin L, Benhamou E, Yoneya M, Furukawa S, Dick F, Slaney M, Chait M (2019) Rapid ocular responses are modulated by bottom-up-driven auditory salience. *J Neurosci* 39:7703–7714.