


# Pupillometry as an Objective Measure of Sustained Attention in Young and Older Listeners

Trends in Hearing  
Volume 23: 1–21  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2331216519887815  
journals.sagepub.com/home/tia  


Sijia Zhao<sup>1</sup>, Gabriela Bury<sup>1</sup>, Alice Milne<sup>1</sup>  and Maria Chait<sup>1</sup> 

## Abstract

The ability to sustain attention on a task-relevant sound source while avoiding distraction from concurrent sounds is fundamental to listening in crowded environments. We aimed to (a) devise an experimental paradigm with which this aspect of listening can be isolated and (b) evaluate the applicability of pupillometry as an objective measure of sustained attention in young and older populations. We designed a paradigm that continuously measured behavioral responses and pupillometry during 25-s trials. Stimuli contained a number of concurrent, spectrally distinct tone streams. On each trial, participants detected gaps in one of the streams while resisting distraction from the others. Behavior demonstrated increasing difficulty with time-on-task and with number/proximity of distractor streams. In young listeners ( $N = 20$ ; aged 18 to 35 years), pupil diameter (on the group and individual level) was dynamically modulated by instantaneous task difficulty: Periods where behavioral performance revealed a strain on sustained attention were accompanied by increased pupil diameter. Only trials on which participants performed successfully were included in the pupillometry analysis so that the observed effects reflect task demands as opposed to failure to attend. In line with existing reports, we observed global changes to pupil dynamics in the older group ( $N = 19$ ; aged 63 to 79 years) including decreased pupil diameter, limited dilation range, and reduced temporal variability. However, despite these changes, older listeners showed similar effects of attentive tracking to those observed in the young listeners. Overall, our results demonstrate that pupillometry can be a reliable and time-sensitive measure of attentive tracking over long durations in both young and (with caveats) older listeners.

## Keywords

attention, hearing, aging, listening effort, auditory scene analysis

Received 15 March 2019; Revised 8 October 2019; accepted 16 October 2019

## Introduction

The ability to sustain attention on a task-relevant stimulus while avoiding distraction from competing information is a fundamental perceptual challenge across sensory modalities. Arguably, this is especially the case in hearing because of the dynamic nature of sound objects. Listening in many natural environments (e.g., a busy train station, a loud restaurant, a noisy classroom) does not only depend on hearing acuity but also on the brain's ability to focus and maintain attention on a specific sound (e.g., an announcement at a train station, a conversation in a restaurant, the teacher's voice in a classroom) while resisting distraction from other concurrent sounds. Understanding "attentive tracking" is central to understanding the challenges faced by the brain during everyday listening and for addressing

impairments in this ability. Indeed, diminished sustained attention capacity is hypothesized to underlie various disorders commonly associated with impaired listening, including auditory processing disorder (e.g., Moore, Ferguson, Edmondson-Jones, Ratib, & Riley, 2010; Moore, Rosen, Bamiou, Campbell, & Sirimanna, 2013), attention deficit hyperactivity disorder (Tucha et al., 2017), autism spectrum disorder (Corbett & Constantine, 2006) and dementia (Berardi, Parasuraman, & Haxby,

<sup>1</sup>Ear Institute, University College London, UK

### Corresponding Author:

Maria Chait, Ear Institute, University College London, 332 Gray's Inn Road, London WC1X 8EE, UK.

Email: m.chait@ucl.ac.uk



2005; Calderon et al., 2001). Failure to maintain attention is also observed in hearing-impaired individuals (Pichora-Fuller et al., 2016) and as a consequence of healthy aging (Mishra, de Villers-Sidani, Merzenich, & Gazzaley, 2014; E. B. Petersen, Wöstmann, Obleser, & Lunner, 2017; Schoof & Rosen, 2014).

To successfully track a given source within a noisy scene, a listener must overcome challenges associated with energetic masking (i.e., extracting the information related to the target from the sound mixture) as well as challenges associated with selecting and continuously following the relevant source from within the background (Shinn-Cunningham & Best, 2008; Woods & McDermott, 2015). Most of the previous work has investigated listening in noisy environments using speech embedded in noise or in a mixture of other speakers, effectively confounding both aspects of tracking. However, it is likely that individual capacity to sustain attention is in itself a factor that will affect listening success. Here, we sought to isolate and continuously monitor this aspect of auditory processing.

A large body of work demonstrates that sustained attention is not static but fluctuates over time and that these behavioral effects are associated with changes in connectivity along a distributed network of brain regions (Fortenbaugh, DeGutis, & Esterman, 2017; Langner & Eickhoff, 2013; Thomson, Besner, & Smilek, 2015). Emerging models postulate that lapses in sustained attention may arise from the weakening of executive processes over time and result in failure to effectively control resource allocation between the main task, distractor suppression, and mind-wandering (Kurzban, Duckworth, Kable, & Myers, 2013). To monitor sustained attention and determine how it is affected by increasing demands on distractor suppression, we designed a paradigm that isolates this facet of listening and measured behavioral and pupillometry responses during 25-s-long trials.

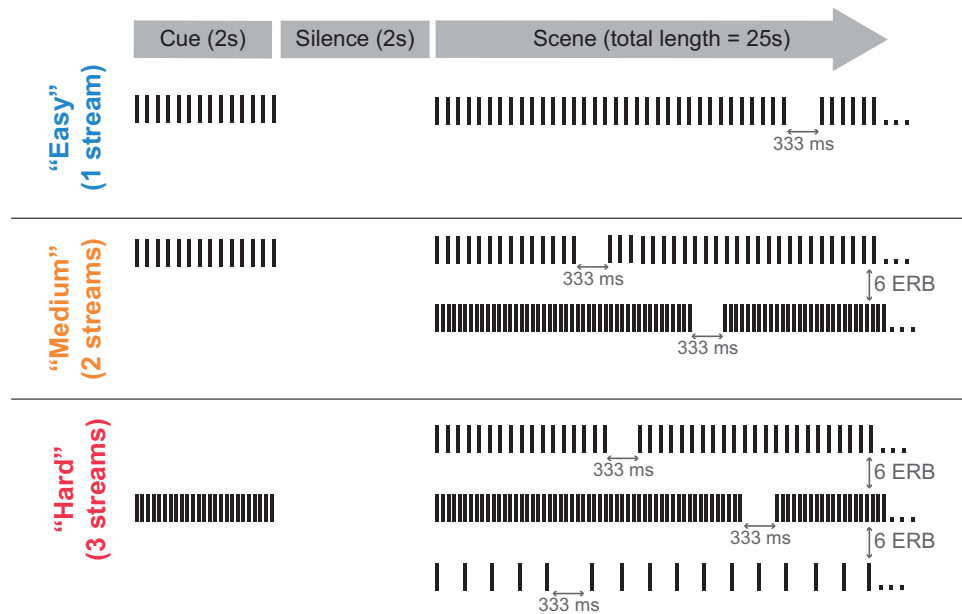
Pupil dilation has long been used as a measure of effort (Beatty, 1982; Bradshaw, 1968; Cabestrero, Crespo, & Quirós, 2009; Granholm & Steinhauer, 2004; Hjortkjær, Märcher-Rørsted, Fuglsang, & Dau, 2018; van der Wel & van Steenbergen, 2018). It is now attracting considerable interest in the auditory modality, because of evidence that pupil dilation can be used as an objective means with which to evaluate challenges to listening (McGarrigle et al., 2014; Peelle, 2018; Pichora-Fuller et al., 2016). The bulk of existing work has used pupillometry to evaluate listening effort associated with degraded or informationally masked speech (Koelewijn, de Kluiver, Shinn-Cunningham, Zekveld, & Kramer, 2015; Koelewijn, Shinn-Cunningham, Zekveld, & Kramer, 2014; Koelewijn, Zekveld, Festen, & Kramer, 2012; Kuchinsky et al., 2014; Naylor, Koelewijn, Zekveld, & Kramer, 2018; Ohlenforst et al.,

2017; C.-A. Wang, Blohm, Huang, Boehnke, & Munoz, 2017; Wendt, Dau, & Hjortkjær, 2016; Wendt, Hietkamp, & Lunner, 2017; M. B. Winn, Edwards, & Litovsky, 2015; M. B. Winn, Wendt, Koelewijn, & Kuchinsky, 2018; Zekveld, Koelewijn, & Kramer, 2018; Zekveld, Kramer, & Festen, 2010, 2011). As a result, these tasks inherently challenged both the ability to cope with a degraded signal and the ability to sustain attention over time. Here, we seek to specifically relate pupil dilation to the challenges of attentive tracking.

There is evidence to suggest that pupil dilation may be particularly correlated to the demands on sustained attention (Hopstaken, van der Linden, Bakker, & Kompier, 2015; Sarter, Givens, & Bruno, 2001). Non-luminance-mediated pupil dilation is at least partially driven by the release of NE (Norepinephrine, also Noradrenaline; Loewenfeld & Lowenstein, 1993) and ACh (Acetylcholine; see recent review Larsen & Waters, 2018). NE release has been consistently linked to arousal and sustained attention through its effects on modulating the response gain of cortical and thalamic neurons (Berridge & Waterhouse, 2003; Sara, 2009). ACh has been associated with activation in the anterior attention system and is hypothesized to play a role in controlling distraction (Berry et al., 2014; Demeter & Sarter, 2013; Kim, Müller, Bohnen, Sarter, & Lustig, 2017; Sarter, Gehring, & Kozak, 2006). We therefore expect that increased demands on sustained attention—including time-on-task and number of distractors—should be revealed in a time-specific manner in the pupil dilation pattern.

The stimuli used in the present experiments are simple artificial “soundscapes” consisting of concurrent, perceptually distinct tone streams (Figure 1) that reduce the demands of segregation and isolate processes associated with object *selection*. Attention is verified and quantified as performance on a gap detection task. Gaps occur in all streams, but listeners are instructed to only respond to those in the target (“Attended”) stream. The scenes are long (~25 s) and the task therefore requires listeners to maintain sustained attention over long durations and actively resist distraction from the other concurrent streams within the scene.

We address two aims: The first aim (Experiment 1) is to ascertain whether pupillometry can be a reliable and time-sensitive measure of the effort associated with attentive tracking over long durations, similar to those over which listeners must maintain attention in ecologically relevant situations. Previous work has used coarse pupil measures (peak dilation) and over relatively short intervals (most investigations have focused on the first 5 s; but see Hjortkjær et al., 2018). In contrast, we sought to measure instantaneous pupil diameter changes over a period of ~25 s. We hypothesized that the harder task conditions will be associated with heightened sustained



**Figure 1.** A schematic representation (not to scale) of the stimuli in Experiment 1. “Scenes” consist of 1 (Easy), 2 (Medium) or 3 (Hard) concurrent perceptually distinct tone streams that model auditory sources. Each source is amplitude modulated at a unique rate to increase distinctiveness. The sources are widely set apart in frequency (6 ERB). On each trial, participants are instructed (via a 2-s-long cue sound) to attend to one of the streams (“target”). Attention is verified and quantified as performance on a gap detection task. Gaps occur in all streams, but listeners are instructed to only respond to those in the target stream. The scenes are long (25 s) and as such the task requires listeners to maintain sustained attention over extended durations and actively resist distraction from the other concurrent streams within the scene. ERB = Equivalent Rectangular Bandwidth.

pupil dilation, reflective of the increased effort to sustain attention.

Our second aim is to determine whether pupillometry as a measure of effort to sustain attention is also applicable to older listeners. Attentive capacity is known to decline with age (e.g., Brosnan et al., 2018; Dørum et al., 2016; Lufi, Segev, Blum, Rosen, & Haimov, 2015; Tu et al., 2018; van der Leeuw et al., 2017). An objective measure of sustained auditory attention would, therefore, be useful to quantify such difficulties and assess intervention outcomes. However, there are known changes to ocular physiology associated with healthy aging (Bitsios, Prettyman, & Szabadi, 1996; Guillon et al., 2016; Tekin et al., 2018; B. Winn, Whitaker, Elliott, & Phillips, 1994) that might limit the efficacy of pupillometry in this population (Piquado, Isaacowitz, & Wingfield, 2010; Van Gerven, Paas, Merriënboer, & Schmidt, 2004). In Experiment 2, we expected older listeners’ pupil dilation patterns to be similar to those observed in the younger group despite possible physiological differences.

## Experiment 1: Young Listeners

### Methods

**Participants.** Thirty-three paid participants (21 females; mean age = 22.9 years, range = 18–31) took part in

this study. All reported normal hearing and no history of neurological disorders. Experimental procedures were approved by the research ethics committee of University College London, and written informed consent was obtained from each participant. Thirteen participants were excluded from the pupillometry analysis due to poor behavioral performance, leaving a subset of 20 participants (14 females, mean age = 22.7 years, range = 18–30 years).

**Stimuli.** Stimuli were 25-s-long artificial acoustic “scenes” that contained 1, 2, or 3 concurrent tone-pip sequences (“streams”). Each stream had a unique carrier frequency and pulse rate. Carrier frequencies were selected from a pool of 18 ERB-spaced (Equivalent Rectangular Bandwidth; Glasberg & Moore, 1990) values between 500 and 4000 Hz with the constraint that the separation between streams (in the 2- and 3-stream condition) was exactly 6 ERBs. Pulse rates were selected from a pool of four values: 3, 7, 13, or 23 Hz. Tone-pip duration was fixed at 30 ms (including 10 ms rise and fall; raised cosine). Together, the unique combination of frequency and pulse rate associated with each stream supported the perception of the scene as consisting of several concurrent, segregable “auditory objects.” To control for perceived loudness, the overall scene intensity was kept

constant across scene-size conditions. As a consequence, individual stream intensity decreased with scene size.

Each stream contained either two or three silent gaps. These were created by removing the appropriate number of tones to generate a silent gap of around 333 ms (the minimum length of a gap in the 3 Hz pulse rate stream). Silent gaps could not occur within the first or last 2 s of a stream sequence or within 2 s of one another (including across streams). Participants were instructed to monitor one of the streams (“target”) for gaps while ignoring gaps in the distractor streams. The target stream was indicated by means of a 2,000 ms cueing tone-pip sequence which preceded each trial. The scene was then presented following a 2,000 ms silent gap (see Figure 1). In the three-stream condition, the target stream was always the middle-frequency stream. To facilitate comparison across conditions, stimuli were created in triplets containing the same target stream across all three conditions. These were then presented in random order during the experimental session.

**Procedure.** Participants sat with their head fixed on a chinrest in front of a monitor (24-in. BENQ XL2420T with a resolution of  $1,920 \times 1,080$  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 while monitoring the cued target stream for gaps. They were to respond (button press) as quickly as possible when a gap was detected while ignoring gaps in the distractor streams. Visual feedback (number of misses and false alarms [FAs]) was presented for 1,500 ms at the end of each trial.

Stimuli were presented in a random order, such that on each trial, the specific condition was unpredictable until scene onset. Sounds were delivered diotically to the participants’ ears with Sennheiser HD558 headphones (Sennheiser, Germany) via a Creative Sound Blaster X-Fi sound card (Creative Technology, Ltd.) at a comfortable listening level self-adjusted by each participant. Stimulus presentation and response recording were controlled with the Psychtoolbox package (Psychophysics Toolbox Version 3; Brainard, 1997) on MATLAB (The MathWorks, Inc.).

The entire experimental session lasted approximately 2 hr. Participants first completed a short practice block followed by six experimental blocks comprised of 12 trials each (~6.5 min, four trials per condition). In total, 72 trials (24 trials per condition) were presented in a random order for each participant.

**Analysis of behavioral data.** Key presses occurring within 0.3 s of a previous key press were considered to be accidental and removed from the analysis. A key press was

classified as a hit if it occurred 0.3 to 1.5 s following a target gap. Hit rate (HR) was computed for each subject, in each condition, as the ratio between detected versus presented gaps in the target stream. All key presses that were not classified as a hit were classified as FAs. These were summed and averaged across trials as a measure of distractibility. Only trials on which participants performed well were included in the pupillometry analysis. “Successful trials” were those where all the target gaps were correctly detected (100% hits) and which included at most one FA. All other trials were classified as “bad trials” and removed from subsequent pupillometry analysis. These three measures, HR, #FA, and #bad trials, are plotted as measures of performance in Figures 2, 3, 6, and 7. Note that FA is quantified as a count (and not as a rate). This is because false responses can happen at any time during the trial.

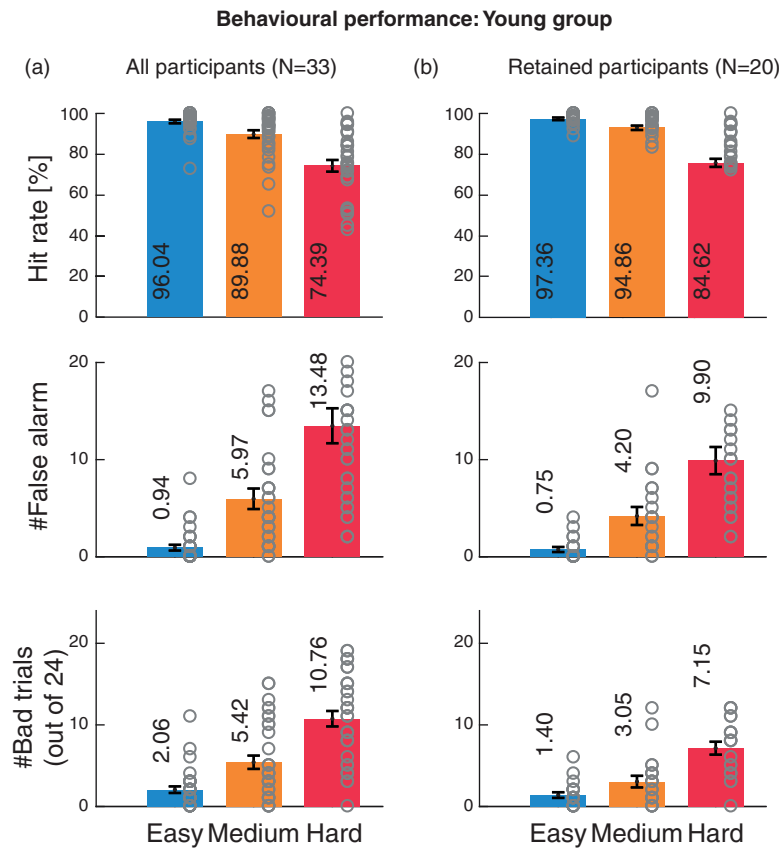
To quantify changes to behavior during the unfolding trial, behavioral data were also analyzed over 5 time bins of 5 s ([0–5] s, [5–10] s, [10–15] s, [15–20] s, and [20–25] s). Mean HR and mean #FAs were computed for each condition in each time bin.

Due to normality-violating ceiling effects in some of the conditions, the behavioral data were analyzed with a nonparametric repeated measures test (Friedman-related measures test). The  $p$  value was a priori set to  $p < .05$ .

**Pupil diameter measurement.** An infrared eye-tracking camera (Eyelink 1000 Desktop Mount, SR Research Ltd.) was positioned at a horizontal distance of 65 cm away from the participant. The standard 5-point calibration procedure for the Eyelink system was conducted prior to 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. Prior to each trial, the eye-tracker automatically checked that the participants’ eyes were open and fixated appropriately; trials would not start unless this was confirmed.

**Analysis: Pupillometry.** As described earlier, only “successful trials” (i.e., those trials on which we can be sure that participants actively tracked the target stream) were included in the pupillometry analysis. To equate the number of trials analyzed per condition, per subject, the number of trials per condition was set to 12 (this number was determined based on the performance of the worst retained participant on the most difficult condition).





**Figure 2.** Behavioral performance of the young group (Experiment 1). Performance measures were: hit rate, number of false alarms, and number of bad trials. (a) Data from all participants ( $N = 33$ ). (b) Data from the participants retained for the pupillometry analysis ( $N = 20$ ). See “Methods” for retention criteria. Gray circles indicate individual data. The task conditions are labeled by difficulty: Easy condition = 1 stream; Medium condition = 2 streams; Hard condition = 3 streams. All performance measures were significantly modulated by task difficulty. Error bar is  $\pm 1$  SEM. As the scene size grows, participants systematically struggle to resist the distraction, detecting fewer targets and making more false alarms. This demonstrates that this task models in a suitable way the competition for processing resources in crowded acoustic scenes.

**Preprocessing.** Only the left eye was analyzed. To measure the pupil dilation response (PDR) associated with tracking the acoustic “scene,” the pupil data from each trial were epoched from 0.5 s prior to “scene” onset to “scene” offset (25-s post-onset). For each trial, baseline correction was applied by subtracting the mean pupil diameter over the pre-onset interval (0.5-s pre-onset).

The data were smoothed with a 150 ms Hanning window and down-sampled to 20 Hz. Intervals where full or partial eye closure was detected (e.g., during blinks) were automatically treated as missing data and recovered using shape-preserving piecewise cubic interpolation. The blink rate was low overall. In both young and older (see later) participant groups and for all conditions, the average blink rate (defined as the proportion of excluded samples due to eye closure) was approximately 5% ( $SD = 5\%$ ). Blinks were distributed evenly over the trial duration. For each participant, the pupil diameter was time-domain-averaged across all epochs of

each condition to produce a single time series per condition. In the main analysis, we focus on absolute pupil diameter change relative to baseline (in mm). All statistics are based on repeated measures comparisons and therefore controlled for intersubject variability (see later). We note that identical results are obtained by  $z$  score normalizing based on pupil size statistics over the pre-onset period (see Figures S1, S2, and S4 in Supplementary Materials).

**Time series statistical analysis.** To identify time intervals in which a given pair of conditions exhibited PDR differences, a non-parametric bootstrap-based statistical analysis was used (Efron & Tibshirani, 1994). The difference in time series between the conditions was computed for each participant, and these time series were subjected to bootstrap resampling (1,000 iterations; with replacement). At each time point, differences were deemed significant if the proportion of bootstrap iterations that fell above or below zero was more than 95%

(i.e.,  $p < .05$ ). Any significant differences in the pre-onset interval would be attributable to noise, and the largest number of consecutive significant samples' pre-onset was used as the threshold for the statistical analysis for the entire epoch.

The main analyses focus on repeated measures comparisons as described earlier. However, we later also compared data (coefficient of variation) between the young and older groups. This was achieved with an "independent samples" bootstrap-based resampling: On each iteration,  $N$  data sets ( $N = 19$  here; based on the number of subjects in the older group) were selected (with replacement) from each group and a difference between means was computed. Further steps were the same as described earlier.

**Participant exclusion criteria.** Participants with more than 50% of bad trials on the hardest condition (three streams) were excluded from the main analysis.

## Results

**Behavioral performance.** Figure 2(a) shows behavioral performance across the full group of participants ( $N = 33$ ). The pattern of performance demonstrates that the task becomes increasingly difficult with the addition of distractor streams to the scene (manifested by reduced HR and increased #FA and #bad trials). This suggests that the paradigm successfully manipulates demands on attentive tracking. Thirteen participants performed poorly on the hardest condition, resulting in an insufficient number of "successful trials". These participants were excluded from further analysis. The fact that 30% of participants are excluded suggests that the task loads resources to the extent that it may deplete them in a large proportion of participants.

Figure 2(b) plots the performance of the 20 retained participants (those who had at least 12 successful trials in the hardest condition). Performance was evaluated with a nonparametric, repeated measures analysis (Friedman-related measures test) with condition (one stream—"Easy," two streams—"Medium," and three streams—"Hard") as factor. All performance measures (HR, #FA, and #bad trials) yielded a main effect of condition: for HR,  $\chi^2 = 25.78$ ,  $p < .001$ ; for #FA,  $\chi^2 = 32.35$ ,  $p < .001$ ; and for #bad trials,  $\chi^2 = 31$ ,  $p < .001$ . Post hoc tests (related samples Wilcoxon signed-rank test) demonstrated significant differences between all conditions for HR (all  $p \leq .026$ ), #FA (all  $p \leq .001$ ), and #bad trials (all  $p \leq .006$ ).

In addition to quantifying the overall effects, we examined how performance evolved over the duration of the trial by separating the trial into 5 s time bins (Figure 3(b)). For HR, a Friedman-related measures test revealed no difference between time bins in the

Easy ( $p = .264$ ) and Medium ( $p = .135$ ) conditions but a significant effect for the Hard condition ( $\chi^2 = 18.88$ ,  $p = .001$ ) consistent with the gradually declining performance observed from Bin 3 onwards. For #FA, the same test revealed no difference between time bins in the Easy ( $p = .139$ ) and Medium ( $p = .082$ ) conditions but a significant effect for the Hard condition ( $\chi^2 = 13.55$ ,  $p = .009$ ) consistent with a peak in FA observed at Bin 3.

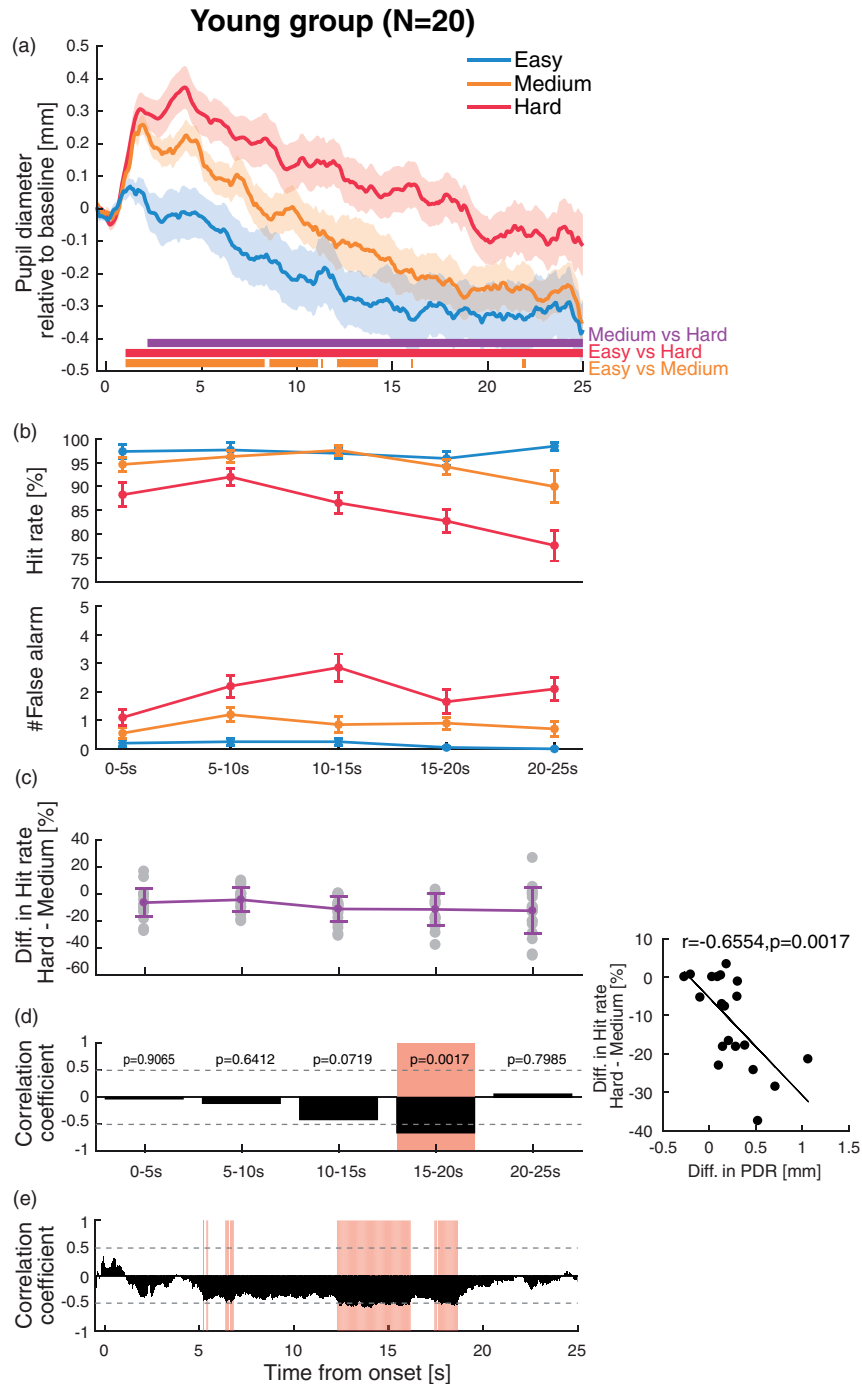
**The PDR as a measure of effort to sustain attention.** Figure 3(a) plots the average pupil diameter data (relative to the pre-onset baseline) as a function of time. Note that the baseline was not taken at a complete resting state but during a brief silent interval (2 s) that occurred between the presentation of the cue and the onset of the scene. At this point, all conditions are equiprobable.

All three conditions share a similar PDR pattern. Immediately after scene onset ( $t = 0$ ), the pupil diameter rapidly increased and reached a peak within 2 s. A significant difference between the PDR to the Easy versus Medium and Hard tracking conditions emerged roughly 1 s after onset. The difference between the Medium and Hard conditions emerged 2.15 s after onset. After the initial peak in the Hard condition (at 2 s), the pupil diameter continuously climbed to a second peak at 4.1 s.

Following the initial dilation, the pupil diameter gradually decreased throughout the epoch but in a manner that preserved the differences between the different conditions. The difference between the Medium and Easy conditions was no longer significant after 14.25 s. However, the PDR to the Hard condition remained considerably above the other two conditions throughout the epoch. Note that the negative pupil diameter values later in the trial reflect the fact that pupil diameter reduced beyond its size during the pre-trial (baseline) period. This likely happens due to the presence of pupil dilation in the pre-trial period, reflecting the anticipation of the onset of the scene (e.g., Bradshaw, 1968; Wierda, van Rijn, Taatgen, & Martens, 2012).

### Correlation between PDR and behavior at an individual level.

To investigate the relationship between pupil dynamics and behavioral performance on an individual subject level, we correlated within each time bin the HR difference between the Hard and Medium conditions (Figure 3(c)) with the corresponding mean PDR difference. The Easy condition was excluded from this analysis because it was associated with little behavioral variability across participants, consistent with ceiling performance. Correlation coefficients (Spearman) are plotted in Figure 3(d). A significant moderate correlation between PDR and HR was observed between 15 and 20 s after trial onset. This timing corresponds to the time window where the HR of the Hard condition



**Figure 3.** The pupil dilation response reflects effort to sustain attention. (a) Pupil dilation results from the young group ( $N = 20$ ). The solid lines represent the average pupil diameter relative to the baseline (500 ms pre-onset) as a function of time. The shaded area shows  $\pm 1$  SEM. Color-coded horizontal lines at graph bottom indicate time intervals where bootstrap statistics confirmed significant differences between each pair of conditions. Qualitatively identical results are obtained with z score normalized data (see Figure S1). (b) Time-binned behavioral performance. Error bars are  $\pm 1$  SEM. (c) Time-binned HR difference between the Hard and Medium conditions. Error bars are  $\pm 1$  standard deviation. Gray dots represent individual data. (d) Correlation between PDR and HR for each time bin. Within each time bin average, the PDR difference between the Hard and Medium conditions was correlated with the corresponding HR difference (as in (c)). Black bars indicate Spearman correlation coefficients at each time bin. Red shaded areas indicate time interval where a significant correlation was observed. Plotted on the right-hand side is the correlation in the 15–20 s time bin. Each dot represents data from a single subject. (e) Correlation between PDR (Hard–Medium condition) and behavioral performance (HR difference between the Hard and Medium conditions) on an individual subject level. Black bars indicate Spearman correlation coefficients at each time point. Red shaded areas indicate time intervals where a significant correlation ( $p < .05$ ; FWE uncorrected) was observed. This analysis was conducted over the entire trial duration with all significant time points indicated. PDR = pupil dilation response.

demonstrated increased divergence relative to the Medium condition (Figure 3(b)).

For a more time-sensitive analysis, we also correlated the instantaneous PDR difference between the Hard and Medium conditions at every time sample (20 Hz) with the mean overall HR difference between these conditions measured for each participant (Figure 3(e)). Correlation coefficients (Spearman) are plotted as black bars in Figure 3(e). Significant time samples (family-wise error (FWE) uncorrected) are marked in red. In line with the time-binned analysis, a significant correlation between instantaneous PDR and HR was found between  $\sim 12$  and  $\sim 19$  s poststream onset.

## Experiment 2: Older Listeners

Overall, the results from Experiment 1 indicate that pupil dilation is a stable and sensitive measure of effort to sustain attention at the group level and that it is associated with individual subject performance. This finding makes PDR a potentially useful objective tool for evaluating attentive tracking ability. Specifically, measuring PDR may be instrumental for quantifying deficits in attentive tracking often exhibited by older populations. However, a potential drawback is the known physiological changes to the pupil that occur during healthy aging; increased demands on accommodation, reduced pupil diameter, and slower responses are commonly observed (Bitsios et al., 1996; Guillon et al., 2016; Tekin et al., 2018). While the physiological underpinnings of these effects are not fully clear (Bitsios et al., 1996), they manifest as relative pupil size rigidity and may reduce the sensitivity of the PDR as a measure of effort.

In Experiment 2a (“Pupilmetrics”), we first replicated these simple changes in our group of older listeners. In Experiment 2b, we then used a paradigm similar to that in Experiment 1 to measure attentive tracking capacity in a group of older listeners.

## Experiment 2a: Pupilmetrics in Young Versus Older Listeners

There are known changes to pupil reactivity with age (Bitsios et al., 1996; Guillon et al., 2016; Tekin et al., 2018; Winn et al., 1994). These include a smaller resting state diameter, a reduced dilation range, and slower velocity of dilation. Here we sought to both replicate these measures and include additional measures of reactivity to brief sounds, as previous reports are mostly focused on reactivity to light flashes. These measures, recorded during passive listening, would later be used to help interpret the attentive tracking data (below).

## Methods

**Participants.** Twenty paid participants aged 60 years or older (14 females, average age = 70.5 years, range = 63–79 years) participated in this experiment. Data from two participants were excluded due to a technical error. Participants were recruited from the U3A (<https://www.u3a.org.uk/>) and therefore represent a sample of high-functioning older individuals. All reported no neurological or existing ophthalmological disorders. Several of the participants reported having successfully undergone cataract surgery 2+ years before this study. Additional inclusion criteria were near-normal-hearing (see “audiometric profile” later) and normal-range performance on an MCI (mild cognitive impairment) screening test (Addenbrooke’s Cognitive Examination—mobile test). Experimental procedures were approved by the research ethics committee of University College London and written informed consent was obtained from each participant.

The young participants group to which the older data are compared comprised the last 18 participants from Experiment 1, mentioned earlier (five females: mean age = 22.4 years, range = 18–31 years).

**Audiometric profile.** Participants were recruited to this experiment based on evidence of near-normal hearing. This was defined as (air-conducted) pure-tone thresholds of 30 dB HL or better at octave frequencies from 0.25 to 4 kHz in both ears. This range was representative of the frequencies used in our stimuli.

**Pupil diameter measurements.** Measurements were conducted during the same session as Experiment 1 (young participants) and Experiment 2 (older participants). Participants completed a 30-s resting state measurement (in silence) before and after the main experiment. They also completed an auditory-evoked PDR measurement which included the presentation of thirty 500 ms harmonic tones ( $f_0 = 200$  Hz; 30 harmonics) with an intersound interval randomized between 6 and 7 s. Participants listened passively to the sounds while pupil responses were recorded. The screen display remained static (identical to that in the main experiment) and participants maintained fixation on a centrally presented black cross.

**Basic pupilmetrics.** The resting pupil diameter was computed as the median value over the 30-s-long resting state trial. *Variability of the pupil diameter* was calculated as one standard deviation over the same period. *Normalized variability* was calculated as variability divided by the corresponding resting pupil diameter. *Pupil response time* was quantified as the *timing and velocity of the PDR* to the onset of a harmonic tone. PDR onset time was quantified by bootstrap resampling



over individual subject data in each group and defined as the first time point from which a significant difference from zero (95% of bootstrap iterations above 0) was sustained for at least 150 ms. *Velocity* was quantified as the peak derivative during the PDR rise time (see Figure 8).

## Results

Figure 4(a) plots the median pupil diameter over a 30-s “resting state” measurement session before and after the main experiment. A repeated measures analysis of variance on pupil size with timing (pre or post the main experiment) as a within-subject measure and age-group as between-subject measure revealed a main effect of age-group,  $F(1, 34) = 22.04$ ,  $p < .001$ , confirming the observation that age is associated with a decreased pupil size (Bitsios et al., 1996; Guillon et al., 2016; Piquado et al., 2010; Tekin et al., 2018; Winn et al., 1994). We also observed a main effect of time,  $F(1, 34) = 10.50$ ,  $p = .003$ , with no interaction, confirming that in both groups, pupil diameter was reduced after the main experiment.

Pupil size fluctuated over time even under constant luminance and without any external stimulation or task. An analysis of pupil normalized variability (standard deviation of pupil size over the 30-s interval; Figure 4(b)) revealed a main effect of age-group,  $F(1, 34) = 55.30$ ,  $p < .001$ , a main effect of time,  $F(1, 34) = 5.50$ ,  $p = .025$ , and an interaction between age and time,  $F(1, 34) = 5.02$ ,  $p = .032$ . Post hoc tests suggested that the source of the interaction was a null effect in the older group ( $p = .831$ ). Overall, these results reveal a smaller resting state pupil size and smaller variability in pupil size in the older participants (see also Winn et al., 1994). The data also demonstrate a decrease in pupil size after the experimental session in both groups. The younger participants additionally exhibited a decrease in the variability of pupil size after the experimental session. The lack of effect in older people may be due to floor effects.

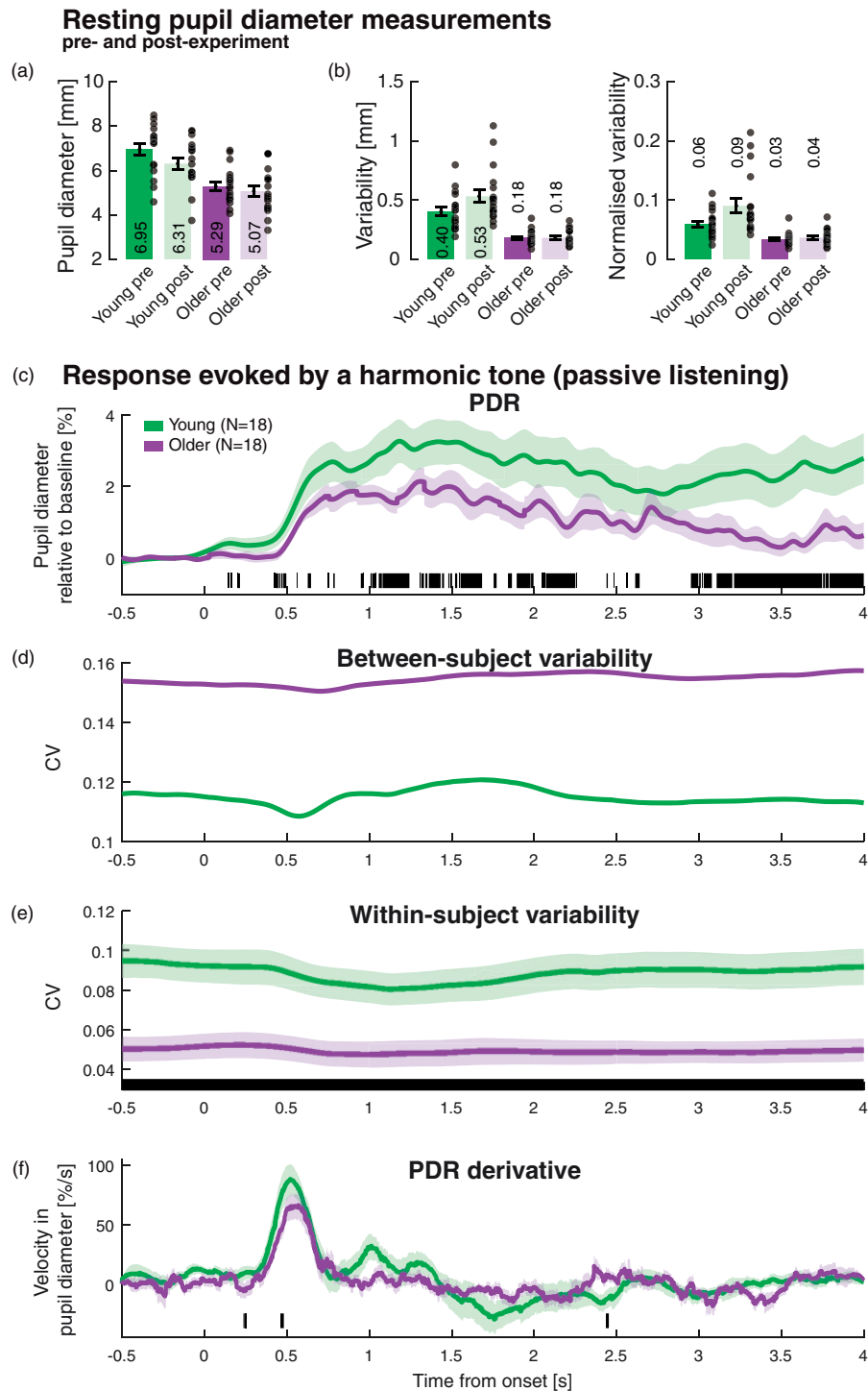
The decrease in pupil size and variability after the experimental session may be a consequence of the effort to maintain fixation during the experimental session or else associated with cognitive fatigue that is linked to the attentional task. We correlated (Spearman) this change in pupil metrics (both absolute size and variability) with behavioral measures (HR and #FA for all three conditions). In both groups, all tests but one (\*) were not significant ( $p > 0.109$ ; \*change in variability correlated with HR in the Easy condition in the young group,  $r = -0.56$ ,  $p = .016$ ). We therefore take this result as indicating (at least for the current N) no evidence for a link between a post-session change in pupil dynamics and individual task-related effort.

An additional simple metric of pupil dynamics is the response to a brief sound event during passive listening. Figure 4(c) plots the harmonic tone evoked PDR in the young and older groups. Both groups exhibited a PDR after the presentation of the tone; the average pupil diameter over the first 3 s following tone onset was significantly above floor as confirmed by a one-sample  $t$  test in each group, young:  $t(17) = 244.46$ ,  $p < .001$  older:  $t(17) = 454.05$ ,  $p < .001$ . The onset of the evoked PDR was 0.46 s in the young group and 0.49 s in the older group. A repeated measures bootstrap confirmed that the PDR of the young group was significantly larger than that of the older group from 0.396 s after onset. We compared the velocity of pupil change between the two groups by taking the first derivative of the PDR (Figure 4(f)). A significant difference between the two groups was observed at  $\sim 0.46$  s post-onset during the rise time of the PDR, suggesting a slower response speed in the older group. This pattern also replicates parallel observations in the context of the darkness-reflex amplitude (Bitsios et al., 1996; Tekin et al., 2018).

For each group, we also analyzed the variability associated with the PDR while controlling for differences in mean pupil diameter. First, we compared the variability across subjects in the young and older groups. To do this, we computed the coefficient of variation (CV=standard deviation/mean) across subjects (Figure 4(d); this analysis was conducted over nonbaseline-corrected data). This revealed a sustained, higher CV in the older group, suggesting that aging is associated with growing individual differences in pupil dynamics. Second, we looked at variability across trials (Figure 4(e)) by calculating the coefficient of variation across trials within each individual and then averaging across subjects for the young and older groups. We found a significantly lower average CV for the older population that was sustained across the epoch (Figure 4(e)). Namely, the older group exhibited substantially lower variability in the PDR response across trials than did the young group.

In sum, older participants exhibited a slower and smaller pupil dilation (pupil diameter change relative to baseline) consistent with reduced reactivity of the pupil in line with previous reports (Piquado et al., 2010; Tekin et al., 2018). The analysis of the sound-evoked PDR also showed a larger between-subject variability and smaller within-subject variability compared with young participants. Since subjects were listening passively, the reduced variability in the older listeners is likely driven by physiological changes to pupil reactivity rather than perceptual engagement per se. The source of these changes may be peripheral (iris physiology) or reflect central deficits in the autonomic system. We will return to this point in the discussion.

In Experiment 2b, we asked whether, despite these age-related changes to pupil response dynamics,



**Figure 4.** Pupil metrics for the young ( $N = 18$ ) and older ( $N = 18$ ) groups. (a) Median pupil diameter computed over a 30 s “resting state” period, pre- and post-main experiment. (b) Variability (standard deviation) in pupil diameter over the resting state measurement. The normalized variability is computed by dividing by the mean diameter. The gray circles indicate individual data. Error bars are  $\pm 1$  SEM. (c) Percentage change in pupil diameter relative to baseline as a function of time from the onset of a brief (500 ms) harmonic tone. Time intervals where bootstrap statistics show significant differences between the means of the two groups are indicated by black horizontal lines. See also Figure S3 for the same analysis on z score normalized data. (d) Between-subject variability in the PDR to harmonic tone across groups. Note that since CV is a single number per time point per group, no cross-group statistics are performed here. (e) Within-subject variability in the PDR to harmonic tone. The solid lines show the average CV across subjects. Error bars are  $\pm 1$  SEM. (f) The derivative of the pupil data shown in (c) as a measure of the velocity of pupil diameter change. Significant differences are indicated as detailed earlier. PDR = pupil dilation response; CV = coefficient of variation.

pupillometry in older listeners can provide a measure of effort to sustain attention.

## Experiment 2b: Attentive Tracking

### Methods

**Participants.** Twenty paid participants aged 60 years or older (those in Experiment 2a; recording during the same session) participated in this experiment. Data from one participant were excluded due to failure to complete the task (>50% bad trials).

**Stimuli and procedure.** The stimulus paradigm (Figure 5) was similar to that in Experiment 1, though we made the task easier. This addressed the concern that the three-stream task (on which 30% of young participants exhibited low performance) may be too difficult for the older listeners who have previously been demonstrated to be more distractible than young controls (Chadick, Zanto, & Gazzaley, 2014; Mishra et al., 2014; Petersen et al., 2017). We therefore chose to limit the scene to two streams. The difficulty was manipulated by varying the spectral separation between streams. The stimulus conditions here included one stream (Easy; identical to Experiment 1), two streams spaced at 10 ERB (Medium), and two streams spaced at 2 ERB (Hard). Note that even in the Hard condition, the spectral separation is such that streams are still perceived as concurrent sources though may be harder to perceptually segregate. Otherwise, all stimulus parameters, generation,

procedure, and analysis were identical to those described for Experiment 1. As in Experiment 1, during the practice session, participants were allowed to adjust the level at which the stimuli were presented to a comfortable loudness. Older listeners tended to choose a higher level than the participants in Experiment 1.

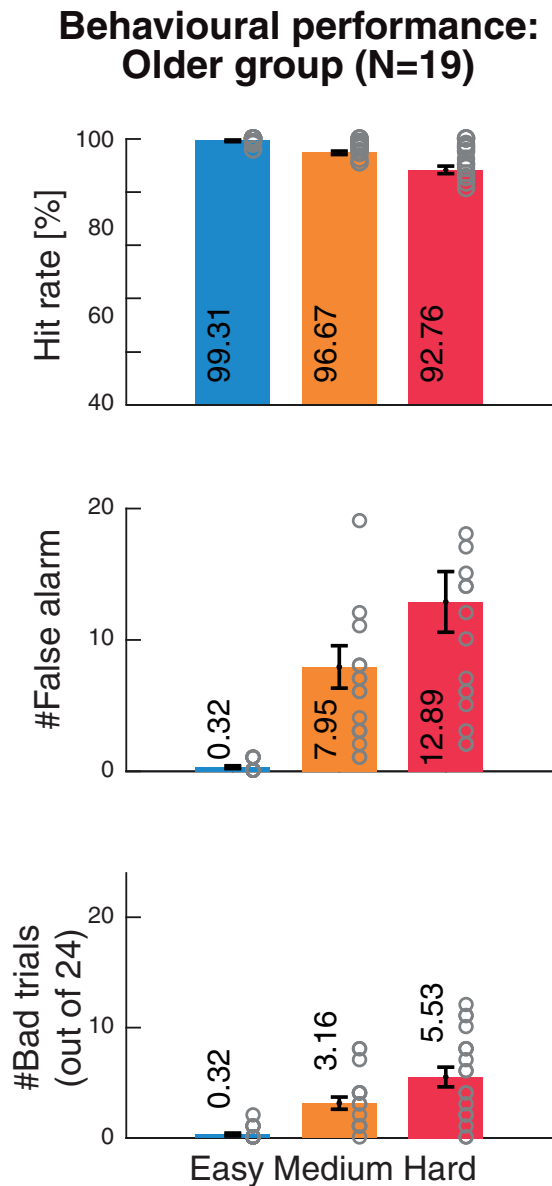
### Results

**Behavioral performance.** Figure 6 shows the behavioral results in the tracking task. Performance was evaluated with a nonparametric, repeated measures analysis (Friedman-related measures test) with condition (1 stream—Easy, two streams—Medium, three streams—Hard) as a factor. All performance measures (HR, #FA, and #bad trials) yielded a main effect of condition: for HR,  $\chi^2 = 17.02$ ,  $p < .001$ ; for #FA,  $\chi^2 = 31.5$ ,  $p < .001$ ; and for #bad trials,  $\chi^2 = 26.4$ ,  $p < .001$ . Post hoc tests (related samples Wilcoxon signed-rank test) demonstrated significant differences between all conditions for HR (all  $p \leq .012$ ), #FA (all  $p \leq .002$ ), and #bad trials (all  $p \leq .001$ ).

In addition to quantifying the overall effects, we examined how performance evolved over the duration of the trial by separating the trial into 5 s time bins (Figure 7(b)). For HR, a Friedman-related measures test revealed no difference between time bins in the Easy ( $p = .665$ ), Medium ( $p = .606$ ), or Hard ( $p = .252$ ) conditions. For #FA, the same test revealed no difference between time bins in the Easy ( $p = .199$ ) and Hard ( $p = .247$ ) conditions, but a significant difference was



**Figure 5.** A schematic representation of the stimuli in Experiment 2. Stimuli were similar to those in Experiment 1, with the exception that difficulty was varied by changing the distance between streams. The Easy condition consisted of a single stream (identical to that in Experiment 1). The Medium condition consisted of two concurrent streams separated by 10 ERB. The Hard condition consisted of 2 concurrent streams separated by 2 ERB. Other parameters are identical to those in Experiment 1. ERB = Equivalent Rectangular Bandwidth.



**Figure 6.** Behavioral performance of the older group (Experiment 2). Performance measures were: average hit rate, number of false alarms, and number of bad trials for retained participants ( $N = 19$ ). Gray circles indicate individual data. All performance measures were significantly modulated by task difficulty. Error bars are  $\pm 1$  SEM.

observed for the Medium condition ( $\chi^2 = 16.52$ ,  $p = .002$ ). This is consistent with the peak in #FA seen during Bin 2 (5–10 s).

HRs were higher than anticipated (and overall higher than those exhibited by the younger group; though note the task for the younger participants was harder). However, FA numbers were equivalent to those exhibited by the younger listeners in Experiment 1, despite the

lower difficulty of the task in Experiment 2. This is consistent with an increased propensity for distraction in older listeners.

*The PDR as a measure of effort to sustain attention.* Figure 7 (a) plots the average pupil diameter data across the older listener group ( $N = 19$ ) as a function of time relative to the pre-onset baseline. As for the young group, the data were baselined relative to the silent interval which preceded the scene onset. Consistent with the observations from Experiment 1 (Figure 3(a)), the older listeners' pupil response also revealed a stable, positive relationship between the amount of effort required to sustain attention during listening and the pupil diameter.

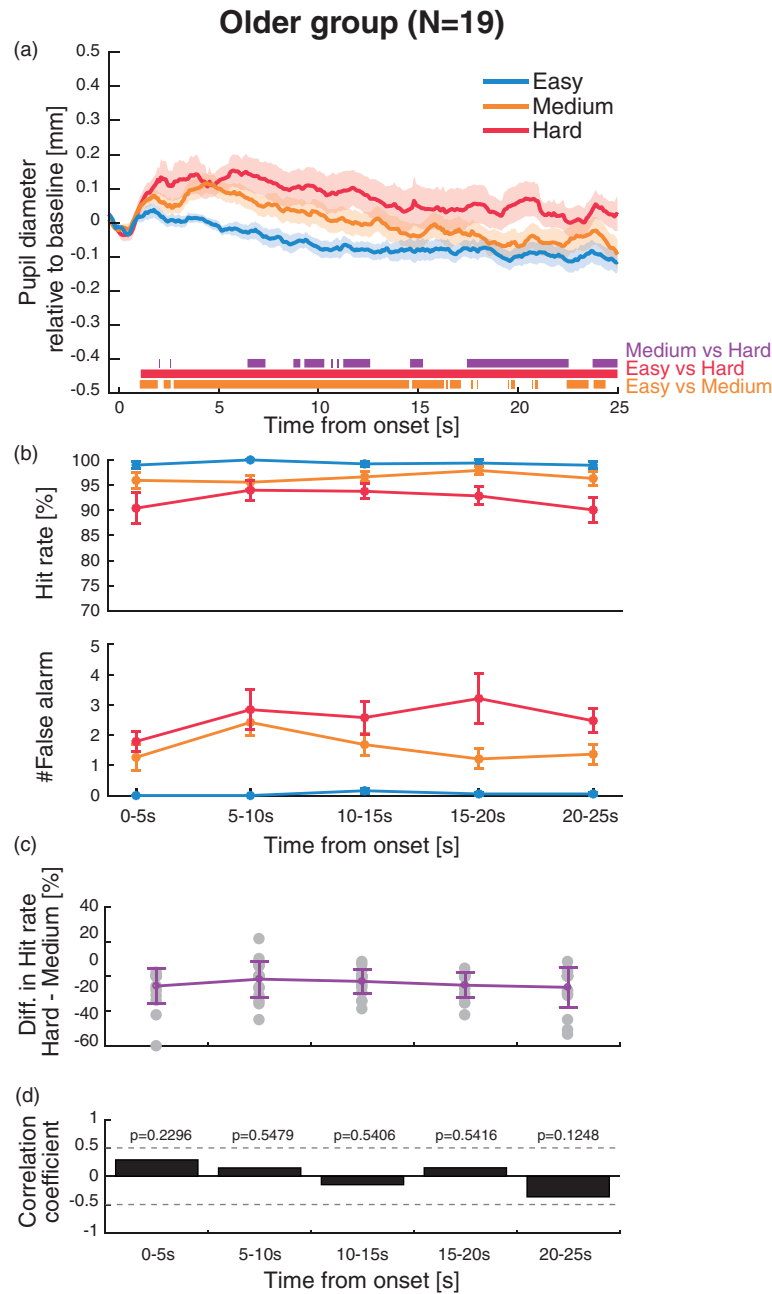
The PDR to the Hard and Medium conditions exceeded the PDR to the Easy condition from 1.1 s post-onset; the PDR to the Hard condition also exceeded the PDR to the Medium condition from 6.4 s. However, unlike for the young group, we failed to find any systematic relationship between the PDR and individual performance (Figure 7(c) and (d)). There could be several reasons for this, including factors associated with task difficulty or lack of sufficient pupil reactivity in the older population.

To explore differences in pupil dynamics between the young and older groups, and specifically to compare response variability, we examined each group's responses to the Easy condition (a single stream). This condition was identical across Experiments 1 and 2b and evoked ceiling performance in both young and older subject groups. Figure 8(a) plots the PDR to the Easy condition in the young and older group. While initially overlapping, responses from the two groups diverged partway through the trial (after about 12 s; a similar result is obtained with  $z$  score normalized data, see Figure S4 in Supplementary Materials). To compare the variability associated with the PDR in each group while controlling for differences in mean pupil diameter, we computed the coefficient of variation ( $CV = \text{standard deviation}/\text{mean}$ ) across subjects (Figure 8(b); this analysis was conducted over nonbaseline-corrected data). Figure 8(b) demonstrates that while the between-subject variability in the older group was relatively stable, that of the young group gradually increased over the trial duration. The CVs diverged from 8-s post-onset until trial offset.

In Figure 8(c), the coefficient of variation was computed for each subject (across trials) and then averaged to produce a measure of within-subject variability. This analysis confirmed that the young group exhibited a relatively larger within-subject variability than the older group, especially following trial onset and after midway through the trial.

Overall, both of these effects demonstrate substantial, time-dependent differences in pupil dynamics between the older and young populations. These differences



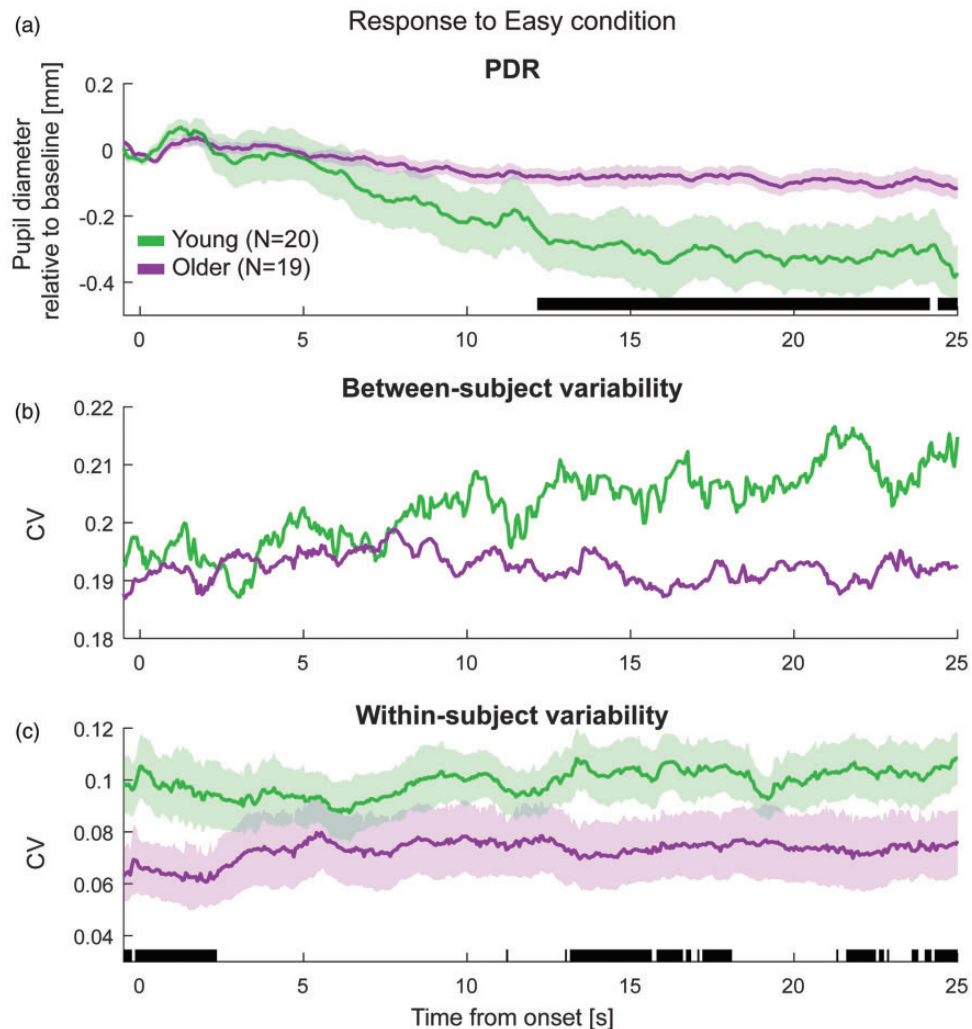


**Figure 7.** The pupil dilation response reflects effort to sustain attention. (a) Pupil dilation results from the older group ( $N = 19$ ). The solid lines represent the average pupil diameter as a function of time relative to the baseline (500 ms pre-onset). The shaded area shows  $\pm 1$  SEM. Color-coded horizontal lines at graph bottom indicate time intervals where bootstrap statistics confirmed significant differences between each pair of conditions. Qualitatively identical results are obtained with z score normalized data. See Figure S2. (b) Time-binned behavioral performance. Error bars are  $\pm 1$  SEM. (c) Time-binned HR difference between the Hard and Medium conditions. Error bars are  $\pm 1$  standard deviation. Gray dots represent individual data. (d) Correlation between PDR and HR for each time bin. Within each time bin average, PDR difference between the Hard and Medium conditions was correlated with the corresponding HR difference (as in (C)). Black bars indicate Spearman correlation coefficients at each time bin. No significant correlations were observed.

mirror those observed (in the absence of a task) in Experiment 2a (Figure 4) and suggest that these differences are attributable to physiological changes to pupil reactivity rather than task engagement. We return to this point in the discussion.

## Discussion

Extending previous research that used pupil dilation as a measure of episodic listening, here we demonstrate that pupillometry can be applied to evaluating sustained auditory attention over long durations that are relevant



**Figure 8.** Comparison of the PDR to the Easy condition in the young and older groups. (a) A comparison of the PDR across groups. The shaded area shows  $\pm 1$  SEM. The black line at graph bottom indicates time intervals where bootstrap resampling confirmed significant differences between groups. Similar effects are also seen in z score normalized data; see Figure S4. (b) Between-subject coefficient of variation (CV) against time. Despite the fact that this condition was identical across groups, the young group exhibited a substantially larger between-subject variability than the older group after 8-s post onset. Note that since CV is a single number per time point per group, no cross-group statistics are performed here. (c) Within-subject coefficient of variation against time—computed by overall trials for each subject. The solid lines present the average CV across subjects. The shaded area shows  $\pm 1$  SEM. The older listeners exhibited relatively smaller across-trial variability, consistent with reduced pupil reactivity. The black horizontal line indicates time intervals where bootstrap resampling confirmed significant differences between groups. PDR = pupil dilation response; CV = coefficient of variation.

to real-life listening. Our results reveal that in young listeners, pupil dilation provides a robust measure of effort to sustain attention. Task difficulty modulated pupil diameter at the group level and revealed modulations of pupil dilation that were correlated with individual subject performance. This opens the possibility of using pupillometry as an objective measure of sustained attention and for characterizing failure of attention in various populations. We provide evidence that similar effects are also obtainable from older listeners but with certain caveats which will be discussed later.

### *Behavioral Measures of Attentive Tracking*

To provide tight control of both stimulus features and the behavioral task, we used simple artificial acoustic “scenes” that allowed us to isolate the demands associated with attentive tracking from other concurrent perceptual challenges. We showed that performance decreased substantially with the number of elements (concurrent streams) in the scene (Experiment 1) and was also modulated by their spectral proximity (Experiment 2b) suggesting that this task is a suitable model with which to capture the challenges of

competition for processing resources in crowded acoustic scenes.

Specifically, the task has several key features: (a) to succeed, listeners must continuously monitor the target stream as even momentary distraction may cause them to miss a target gap; (b) listeners are required to respond to multiple events within the unfolding sequence, providing precise tracking of attention; and (c) the task is devoid of memory and semantic confounds commonly associated with speech stimuli, avoiding interactions that may arise as a consequence of the depletion of resources (Mattys & Wiget, 2011; Schmidt, Scharenborg, & Janse, 2015). The use of simple sounds (not speech) also circumvents many practical issues including those related to language proficiency, making the paradigm appropriate for a variety of subjects from children to older listeners.

In future work, the stimuli can be made increasingly complex by varying scene size, source trajectories (e.g., introducing frequency modulation), spatial extent, and so forth. Due to their narrowband nature, the signals can also be easily adjusted to fit the hearing profile of the individual tested.

Because of our policy of only including successful trials in the pupillometry analysis, we had to exclude 30% of the young participants who failed to achieve a sufficient number of trials for analysis. That about a third of our cohort failed on the hardest condition suggests that resources were likely exhausted by the task. Whether there are any cognitive markers which might differentiate those who succeeded from those who failed is an interesting question for future work.

### *Pupil Measures in Young Listeners Track Effort to Sustain Attention*

Manipulation of effort through varying task difficulty is intrinsically associated with reduced performance, that is, an increasing number of trials on which participants fail to accomplish the task. It is common practice in the field to analyse all trials, irrespective of their behavioral outcome (e.g., Koelewijn et al., 2012, 2014, 2015; Kuchinsky et al., 2014; Naylor et al., 2018; Ohlenforst et al., 2017; Wang et al., 2017; Wendt et al., 2016, 2017; Winn et al., 2015, 2018; Zekveld et al., 2010, 2011, 2018). In contrast, here we chose to focus on correct trials only.

While errors may occur despite participants being fully focused on the task, they may also arise from various other sources including inadvertent disengagement from the task or mind-wandering (Fortenbaugh et al., 2017). Because we have no evidence for the underlying cognitive process that resulted in the error, there is therefore a risk, which is further exacerbated by the long durations of trials in the present paradigm, that pupil activity measured during “failed” trials may be contaminated by processes

linked with the failure of attention and/or disengagement (Hopstaken et al., 2015; van den Brink, Murphy, & Nieuwenhuis, 2016; Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013; Pelagatti, Binda, & Vannucci, 2018; Smallwood, Fishman, & Schooler, 2007).

To address this concern, we implemented a policy of only including successful trials in the pupillometry analysis. These are defined as trials on which all target gaps (either 2 or 3) have been correctly identified and where the participant had at most one FA. In this way, we focused on trials where resources were appropriately allocated and distractors successfully ignored. Therefore, any differences observed between conditions reveal pure effects of task demands without contamination from other cognitive processes such as those related to task disengagement. Furthermore, trials of different difficulty were presented in an intermixed order so as to control overall task difficulty effects on baseline pupil activity.

Adopting these criteria, we found that challenging stream tracking conditions were accompanied by large, sustained pupil dilation that mirrored behavioral performance on the group and individual level: Listeners who found the task harder exhibited bigger changes in pupil size.

Despite using time-constant stimulus parameters, the behavioral data indicated that task difficulty was not stable but increased partway through the trial. Statistical analysis showed that this particularly affected the hardest condition, where the HR decreased substantially above the easier conditions from about 10 s onward. The same pattern was present in the pupillometry data. Notably, it was around this time that significant individual-level correlations between pupil diameter and performance were observed. These effects are consistent with multiple observations that the ability to sustain attention deteriorates with time-on-task (Fortenbaugh et al., 2017; Thomson et al., 2015) and is hypothesized to reflect weakened control of cognitive resources (Berry, Sarter, & Lustig, 2017; Esterman, Reagan, Liu, Turner, & DeGutis, 2014; Pattyn, Neyt, Henderickx, & Soetens, 2008; Sarter & Paolone, 2011; Thomson et al., 2015). Indeed, the FA number peaked mid-trial, reflecting the fact that participants were increasingly unable to resist distraction from the nontarget streams. This is in line with previous proposals that reduced resource control is associated with impaired distractor filtering (Sarter et al., 2001).

It is interesting that significant correlations were observed partway through the trial but did not persist until offset, despite the fact that HR showed a consistent deterioration. A possible explanation for this effect is that expectation of trial offset affects pupil dynamics in a manner that interferes with the correlation with behavior.

### *Is Pupillometry in Older Listeners a Useful Objective Measure?*

Aging is associated with loss of function within the peripheral auditory system that leads to a broad range of auditory processing impairments. In addition, normal aging is associated with various deficits of cognitive, executive, and sustained attention function that have expansive perceptual consequences across sensory modalities. In the context of hearing, these deficits may have wide-ranging implications for listening in crowded environments, such as the ability to attend to a relevant sound source and avoid distraction by concurrent sounds. Routine audiological assessments are not sensitive to these impairments, resulting in suboptimal understanding and management of these conditions. Pupillometry may be a promising tool to quantify such impairments as it is cheap, portable, and noninvasive.

Previous work raised the concern that known age-related changes to ocular physiology may limit the utility of “cognitive pupillometry” in this population. Notably, aging is commonly associated with increased rigidity of the pupil (senile miosis; Meller, 1904) which results in overall decreases in pupil size, range of pupil dilation and response speed (Bitsios et al., 1996; Tekin et al., 2018; also replicated here in Figure 4(c)). The restricted range of the pupil in older listeners may thus limit the ability to observe small, cognitive-state mediated changes to pupil size (Piquado et al., 2010; Van Gerven et al., 2004). However, here we observed significant sustained effects despite quite small behavioral differences between conditions, suggesting that pupillometry can be a sensitive measure of effort to sustain attention in an older population.

Specifically, we demonstrate clear and robust effects of task difficulty on pupil diameter in our group of normal hearing, high performing older individuals. These effects were sustained over the trial duration and paralleled group-level behavioral performance (Figure 7). However, they also differed from those observed in the young group in several important respects: First, unlike in the young group, we did not see any correlation with individual performance. This may be because the task was too easy. Although we decided on the present task based on pilot experiments, the resulting performance was better than expected. Future work should adjust the difficulty to each listener independently. Second, the pupil data from older participants exhibited substantially smaller variability across participants, trials and time, even when accounting for the smaller baseline pupil size (Figure 8).

To control for the difference in pupil size range between older and younger listeners, Piquado et al. (2010) adopted an approach where the pupil data were normalized by the absolute difference in pupil diameter

measured in bright versus dark lighting conditions. Processing-effort-related PDR was then expressed as a proportion of the dynamic range. This approach is based on the premise that pupil reactivity to changes in lighting is similar to that associated with central neuromodulatory processes. Not enough is understood about the underlying circuitry to assess the validity of this assumption. Here, we chose not to normalize pupil data. Since the statistical analysis is based on within-subject comparisons, the results are not affected by differing pupil range between groups. The analysis of variability further demonstrates that a major source of difference between the two groups is not only pupil size but also within-subject (across trial) mean-corrected variability which is substantially larger in the younger group. This difference is measurable even under passive listening conditions, that is, is not driven by task engagement.

One possibility is that the variability in pupil size present in young listeners may reflect nonstationary physiological noise. The lack of such variability in the older group may thus be taken as an advantage in the sense that it results in a cleaner task-locked signal. However, it is increasingly understood that instantaneous fluctuations in pupil size reflect momentary changes in perceptual state that contribute in important ways to behavioral variability (Allen et al., 2016; Fontanini & Katz, 2008; Kelly, Uddin, Biswal, Castellanos, & Milham, 2008). The reduced variability of the pupil in older populations may make us blind to many of these effects.

Another not mutually exclusive possibility relates to the mechanisms that support pupil dynamics. As we discuss further later, both sympathetic and parasympathetic systems can affect pupil dilation. It is feasible that the decreased variability of the pupil in older participants may be related to a reduction in sympathetic activity (Bitsios et al., 1996; see also Mather & Harley, 2015), while the observed modulation of the PDR as a function of tracking difficulty is produced by the relatively preserved parasympathetic activity.

### *Neuromodulator Effects on Sustained Attention*

Mounting evidence from electrophysiology in animal models has revealed a strong correlation between pupil-size dynamics and activity of NE (Joshi, Li, Kalwani, & Gold, 2016; Phillips, Szabadi, & Bradshaw, 2000; Rajkowski, Kubiak, & Aston-Jones, 1993) and ACh expressing neurons (Reimer et al., 2016; Zaborszky et al., 2015). Pupil size is modulated by the balance between dilator and sphincter muscles in the iris. The dilator muscle is innervated by the sympathetic system which acts by releasing NE, and the sphincter muscle is innervated by the parasympathetic system for which ACh is the major neurotransmitter



(Loewenfeld & Lowenstein, 1993; Steinhauer & Hakerem, 1992; Steinhauer, Siegle, Condray, & Pless, 2004). ACh exerts an inhibitory effect in the oculomotor nucleus of the brain stem leading to relaxation of the sphincter muscles, and therefore also to pupil dilation. Consequently, increased release of NE and ACh both contribute to pupil dilation (Larsen & Waters, 2018); however, whether the effects are independent or also synergistic remains unknown.

NE and ACh are hypothesized to play key roles in supporting cognitive effort and executive control (Aston-Jones & Cohen, 2005; Botvinick, Braver, Barch, Carter, & Cohen, 2001; Sarter et al., 2006; Steinhauer et al., 2004). Specifically, a large body of work has linked NE release to increased arousal (see review Berridge & Waterhouse, 2003) and sustained attention (e.g., Aston-Jones, Rajkowski, Kubiak, & Alexinsky, 1994; Carli, Robbins, Evenden, & Everitt, 1983; Sara, 2009). ACh has been associated among other things with activation in the anterior attention system (which underlies effortful, top-down control of goal-directed behavior; Petersen & Posner, 2012) and is hypothesized to play a role in controlling distraction (Berry et al., 2014; Demeter & Sarter, 2013; Himmelheber, Sarter, & Bruno, 2000; Kim et al., 2017; Sarter et al., 2006). In the context of the present task, it is possible that the observed pupil dilation effects reflect a combination of NE-mediated heightened vigilance as well as ACh-mediated processes linked to the need to maintain focus on the target sequence and avoid distraction from the concurrent, nontarget auditory streams.

Based on the pupillary response pattern observed in their experiments, Bitsios et al. (1996; see also Tekin et al., 2018) argued that the altered pupil dynamics commonly observed in aging subjects and also replicated for the present cohort (Figure 4) are of a central origin and predominantly driven by weakened signaling from the sympathetic system (see also Mather & Harley, 2015).

It is therefore tempting to postulate that the reduced variability in pupil diameter observed here in older listeners (Experiment 2a and b) may reflect the decline in NE-mediated pupil dilation while the preserved effect of attention on average pupil size may be driven by ACh-linked pupil dynamics. This is also consistent with a key role for ACh in supporting attentive listening by suppressing distractors—a main feature of the present task (Berry et al., 2014; Demeter & Sarter, 2013; Himmelheber et al., 2000; Kim et al., 2017; Sarter et al., 2006). Future work with more sensitive techniques in animal models or pharmacological manipulations in humans (see also Steinhauer et al., 2004; Wang et al., 2016; Wang et al., 2018) is needed to tease apart the contribution of ACh and NE to attentive listening.

## Conclusions

The reported experiments demonstrate that pupillometry can be a reliable and time-sensitive measure of the effort associated with sustained listening, extending the use of pupillometry to longer listening tasks beyond the 1 to 5 s stimuli that have been typically used in auditory listening effort research. Our main findings are that in young listeners, pupil dilation correlates with performance such that listeners who experience more difficulty in sustaining attention on the target stream also produce larger pupil dilations. This opens the possibility of using these methods to evaluate listening difficulty, in real time and on the individual level. We further show that similar effects are obtainable in an older population. However, the altered pupil dynamics in that population result in decreased pupil dilation range and slower response speed and may limit observable effects.

## Acknowledgments

Initial pilot experiments were performed by Youna Cho. The authors are grateful to Mathilde Le Gal de Kerangal for helping recruit and screen the older listeners in Experiment 2.

## Data Availability Statement

All raw and processed data are available at [<https://doi.org/10.5522/04/10247612>].

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by funding from the BRC and the EC (Horizon2020 grant to M. C.). Salary support for A.M. also provided by Wellcome Trust [213686/Z/18/Z].

## ORCID iDs

Alice Milne  <https://orcid.org/0000-0002-2000-2842>  
 Maria Chait  <https://orcid.org/0000-0002-7808-3593>

## Supplemental Material

Supplemental material for this article is available online.

## References

- Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *eLife*, 5, e18103. doi:10.7554/eLife.18103
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain

- and optimal performance. *Annual Review of Neuroscience*, 28(1), 403–450. doi:10.1146/annurev.neuro.28.061604.135709
- Aston-Jones, G., Rajkowski, J., Kubiak, P., & Alexinsky, T. (1994). Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. *Journal of Neuroscience*, 14(7), 4467–4480.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276–292. doi:10.1037/0033-2909.91.2.276
- Berardi, A. M., Parasuraman, R., & Haxby, J. V. (2005). Sustained attention in mild Alzheimer's disease. *Developmental Neuropsychology*, 28(1), 507–537. doi:10.1207/s15326942dn2801\_4
- Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, 42(1), 33–84. doi:10.1016/S0165-0173(03)00143-7
- Berry, A. S., Demeter, E., Sabhapathy, S., English, B. A., Blakely, R. D., Sarter, M., & Lustig, C. (2014). Disposed to distraction: Genetic variation in the cholinergic system influences distractibility but not time-on-task effects. *Journal of Cognitive Neuroscience*, 26(9), 1981–1991. doi:10.1162/jocn\_a\_00607
- Berry, A. S., Sarter, M., & Lustig, C. (2017). Distinct frontoparietal networks underlying attentional effort and cognitive control. *Journal of Cognitive Neuroscience*, 29(7), 1212–1225. doi:10.1162/jocn\_a\_01112
- 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 and Ageing*, 25(6), 432–438. doi:10.1093/ageing/25.6.432
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652. doi:10.1037/0033-295X.108.3.624
- Bradshaw, J. L. (1968). Load and pupillary changes in continuous processing tasks. *British Journal of Psychology*, 59(3), 265–271. doi:10.1111/j.2044-8295.1968.tb01139.x
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Brosnan, M. B., Arvaneh, M., Harty, S., Maguire, T., O'Connell, R., Robertson, I. H., & Dockree, P. M. (2018). Prefrontal modulation of visual processing and sustained attention in aging, a tDCS–EEG coregistration approach. *Journal of Cognitive Neuroscience*, 30(11), 1630–1645. doi:10.1162/jocn\_a\_01307
- Cabestrero, R., Crespo, A., & Quirós, P. (2009). Pupillary dilation as an index of task demands. *Perceptual and Motor Skills*, 109(3), 664–678. doi:10.2466/pms.109.3.664-678
- Calderon, J., Perry, R. J., Erzincinoglu, S. W., Berrios, G. E., Dening, T. R., & Hodges, J. R. (2001). Perception, attention, and working memory are disproportionately impaired in dementia with Lewy bodies compared with Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 70(2), 157–164. doi:10.1136/jnnp.70.2.157
- Carli, M., Robbins, T. W., Evenden, J. L., & Everitt, B. J. (1983). Effects of lesions to ascending noradrenergic neurones on performance of a 5-choice serial reaction task in rats; implications for theories of dorsal noradrenergic bundle function based on selective attention and arousal. *Behavioural Brain Research*, 9(3), 361–380. doi:10.1016/0166-4328(83)90138-9
- Chadick, J. Z., Zanto, T. P., & Gazzaley, A. (2014). Structural and functional differences in medial prefrontal cortex underlie distractibility and suppression deficits in ageing. *Nature Communications*, 5, 4223. doi:10.1038/ncomms5223
- Corbett, B. A., & Constantine, L. J. (2006). Autism and attention deficit hyperactivity disorder: Assessing attention and response control with the integrated visual and auditory continuous performance test. *Child Neuropsychology*, 12(4–5), 335–348. doi:10.1080/09297040500350938
- Demeter, E., & Sarter, M. (2013). Leveraging the cortical cholinergic system to enhance attention. *Neuropharmacology*, 64, 294–304. doi:10.1016/j.neuropharm.2012.06.060
- Dorum, E. S., Alnæs, D., Kaufmann, T., Richard, G., Lund, M. J., Tønnesen, S., ... Westlye, L. T. (2016). Age-related differences in brain network activation and co-activation during multiple object tracking. *Brain and Behavior*, 6(11), e00533. doi:10.1002/brb3.533
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Boca Raton, FL: CRC Press.
- Esterman, M., Reagan, A., Liu, G., Turner, C., & DeGutis, J. (2014). Reward reveals dissociable aspects of sustained attention. *Journal of Experimental Psychology: General*, 143(6), 2287–2295. doi:10.1037/xge0000019
- Fontanini, A., & Katz, D. B. (2008). Behavioral states, network states, and sensory response variability. *Journal of Neurophysiology*, 100(3), 1160–1168. doi:10.1152/jn.90592.2008
- Fortenbaugh, F. C., DeGutis, J., & Esterman, M. (2017). Recent theoretical, neural, and clinical advances in sustained attention research. *Annals of the New York Academy of Sciences*, 1396(1), 70–91. doi:10.1111/nyas.13318
- Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). Window to the wandering mind: Pupillometry of spontaneous thought while reading. *The Quarterly Journal of Experimental Psychology*, 66(12), 2289–2294. doi:10.1080/17470218.2013.858170
- Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1–2), 103–138. doi:10.1016/0378-5955(90)90170-T
- Granholm, E., & Steinhauer, S. R. (2004). Pupillometric measures of cognitive and emotional processes. *International Journal of Psychophysiology*, 52(1), 1–6. doi:10.1016/j.ijpsycho.2003.12.001
- Guillon, M., Dumbleton, K., Theodoratos, P., Gobbe, M., Wooley, C. B., & Moody, K. (2016). The effects of age, refractive status, and luminance on pupil size. *Optometry and Vision Science*, 93(9), 1093–1100. doi:10.1097/OPX.0000000000000893
- Himmelheber, A. M., Sarter, M., & Bruno, J. P. (2000). Increases in cortical acetylcholine release during sustained attention performance in rats. *Cognitive Brain Research*, 9(3), 313–325. doi:10.1016/S0926-6410(00)00012-4
- Hjortkjær, J., Märcher-Rørsted, J., Fuglsang, S. A., & Dau, T. (2018). Cortical oscillations and entrainment in speech

- processing during working memory load. *European Journal of Neuroscience*, 0(0). doi:10.1111/ejn.13855
- Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology*, 110, 100–106. doi:10.1016/j.biopsycho.2015.06.013
- Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, 89(1), 221–234. doi:10.1016/j.neuron.2015.11.028
- Kelly, A. M. C., Uddin, L. Q., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2008). Competition between functional brain networks mediates behavioral variability. *NeuroImage*, 39(1), 527–537. doi:10.1016/j.neuroimage.2007.08.008
- Kim, K., Müller, M. L. T. M., Bohnen, N. I., Sarter, M., & Lustig, C. (2017). The cortical cholinergic system contributes to the top-down control of distraction: Evidence from patients with Parkinson's disease. *NeuroImage*. doi:10.1016/j.neuroimage.2017.12.012
- Koelwijn, 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. doi:10.1016/j.heares.2015.02.004
- Koelwijn, 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. doi:10.1016/j.heares.2014.03.010
- Koelwijn, 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(2), 291. doi:10.1097/AUD.0b013e3182310019
- Kuchinsky, S. E., Ahlstrom, J. B., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2014). Speech-perception training for older adults with hearing loss impacts word recognition and effort: Changes in word recognition and effort. *Psychophysiology*, 51(10), 1046–1057. doi:10.1111/psyp.12242
- 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(6), 661–679. doi:10.1017/S0140525X12003196
- Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, 139(4), 870–900. doi:10.1037/a0030694
- Larsen, R. S., & Waters, J. (2018). Neuromodulatory correlates of pupil dilation. *Frontiers in Neural Circuits*, 12. doi:10.3389/fncir.2018.00021
- Loewenfeld, I. E., & Lowenstein, O. (1993). *The pupil: Anatomy, physiology, and clinical applications*. Oxford, England: Butterworth-Heinemann.
- Lufi, D., Segev, S., Blum, A., Rosen, T., & Haimov, I. (2015). The effect of age on attention level: A comparison of two age groups. *The International Journal of Aging and Human Development*, 81(3), 176–188. doi:10.1177/0091415015614953
- Mather, M., & Harley, C. W. (2016). The Locus Coeruleus: Essential for Maintaining Cognitive Function and the Aging Brain. *Trends in Cognitive Sciences*, 20(3), 214–226. https://doi.org/10.1016/j.tics.2016.01.001
- Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65(2), 145–160. doi:10.1016/j.jml.2011.04.004
- McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal of Audiology*, 53(7), 433–445. doi:10.3109/14992027.2014.890296
- Meller, J. (1904). Über hyaline degeneration des Papillarrandes. *Von Graefes Arch Ophthalm*, 59, 221–228.
- Mishra, J., de Villers-Sidani, E., Merzenich, M., & Gazzaley, A. (2014). Adaptive training diminishes distractibility in aging across species. *Neuron*, 84(5), 1091–1103. doi:10.1016/j.neuron.2014.10.034
- Moore, D. R., Ferguson, M. A., Edmondson-Jones, A. M., Ratib, S., & Riley, A. (2010). Nature of auditory processing disorder in children. *Pediatrics*, 126(2), e382–e390. doi:10.1542/peds.2009-2826
- Moore, D. R., Rosen, S., Bamiou, D.-E., Campbell, N. G., & Sirimanna, T. (2013). Evolving concepts of developmental auditory processing disorder (APD): A British Society of Audiology APD special interest group 'white paper'. *International Journal of Audiology*, 52(1), 3–13. doi:10.3109/14992027.2012.723143
- Naylor, G., Koelwijn, T., Zekveld, A. A., & Kramer, S. E. (2018). The application of pupillometry in hearing science to assess listening effort. *Trends in Hearing*, 22. doi:10.1177/2331216518799437
- Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., ... Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research*, 351, 68–79. doi:10.1016/j.heares.2017.05.012
- Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue? *Physiology & Behavior*, 93(1), 369–378. doi:10.1016/j.physbeh.2007.09.016
- Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, 39(2), 204. doi:10.1097/AUD.0000000000000494
- Pelagatti, C., Binda, P., & Vannucci, M. (2018). Tracking the dynamics of mind wandering: Insights from pupillometry. *Journal of Cognition*, 1(1), 38. doi:10.5334/joc.41
- Petersen, E. B., Wöstmann, M., Obleser, J., & Lunner, T. (2017). Neural tracking of attended versus ignored speech is differentially affected by hearing loss. *Journal of Neurophysiology*, 117(1), 18–27. doi:10.1152/jn.00527.2016
- Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 Years after. *Annual Review of Neuroscience*, 35(1), 73–89. doi:10.1146/annurev-neuro-062111-150525
- Phillips, M. A., Szabadi, E., & Bradshaw, C. M. (2000). Comparison of the effects of clonidine and yohimbine on



- pupillary diameter at different illumination levels. *British Journal of Clinical Pharmacology*, 50(1), 65–68. doi:10.1046/j.1365-2125.2000.00225.x
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., ... Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, 37, 5S. doi:10.1097/AUD.0000000000000312
- Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, 47(3), 560–569. doi:10.1111/j.1469-8986.2009.00947.x
- Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1993). Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention. *Society for Neuroscience Abstracts*, (19), 874.
- Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolia, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, 7, 13289. doi:10.1038/ncomms13289
- Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, 10(3), 211–223. doi:10.1038/nrn2573
- Sarter, M., Gehring, W. J., & Kozak, R. (2006). More attention must be paid: The neurobiology of attentional effort. *Brain Research Reviews*, 51(2), 145–160. doi:10.1016/j.brainresrev.2005.11.002
- Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: Where top-down meets bottom-up. *Brain Research Reviews*, 35(2), 146–160. doi:10.1016/S0165-0173(01)00044-3
- Sarter, M., & Paolone, G. (2011). Deficits in attentional control: Cholinergic mechanisms and circuitry-based treatment approaches. *Behavioral Neuroscience*, 125(6), 825–835. doi:10.1037/a0026227
- Schmidt, J., Scharenborg, O., & Janse, E. (2015, August). *Semantic processing of spoken words under cognitive load in older listeners*. Paper presented at Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015), Glasgow, Scotland.
- Schoof, T., & Rosen, S. (2014). The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Frontiers in Aging Neuroscience*, 6, 1–14. doi:10.3389/fnagi.2014.00307
- Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification*, 12(4), 283–299. doi:10.1177/1084713808325306
- Smallwood, J., Fishman, D. J., & Schooler, J. W. (2007). Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic Bulletin & Review*, 14(2), 230–236. doi:10.3758/BF03194057
- Steinhauer, S. R., & Hakerem, G. (1992). The pupillary response in cognitive psychophysiology and schizophrenia. *Annals of the New York Academy of Sciences*, 658(1), 182–204. doi:10.1111/j.1749-6632.1992.tb22845.x
- Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology*, 52(1), 77–86. doi:10.1016/j.ijpsycho.2003.12.005
- Tekin, K., Sekeroglu, M. A., Kiziltoprak, H., Doguizi, S., Inanc, M., & Yilmazbas, P. (2018). Static and dynamic pupillometry data of healthy individuals. *Clinical and Experimental Optometry*, 101(5), 659–665. doi:10.1111/cxo.12659
- Thomson, D. R., Besner, D., & Smilek, D. (2015). A resource-control account of sustained attention: Evidence from mind-wandering and vigilance paradigms. *Perspectives on Psychological Science*, 10(1), 82–96. doi:10.1177/1745691614556681
- Tu, M.-C., Lo, C.-P., Huang, C.-F., Huang, W.-H., Deng, J. F., & Hsu, Y.-H. (2018). Visual attention performances and related cerebral microstructural integrity among subjects with subjective cognitive decline and mild cognitive impairment. *Frontiers in Aging Neuroscience*, 10. doi:10.3389/fnagi.2018.00268
- Tucha, L., Fuermaier, A. B. M., Koerts, J., Buggenthin, R., Aschenbrenner, S., Weisbrod, M., ... Tucha, O. (2017). Sustained attention in adult ADHD: Time-on-task effects of various measures of attention. *Journal of Neural Transmission*, 124(Suppl 1), 39–53. doi:10.1007/s00702-015-1426-0
- Wang, Y., Kramer, S. E., Wendt, D., Naylor, G., Lunner, T., & Zekveld, A. A. (2018). The Pupil Dilation Response During Speech Perception in Dark and Light: The Involvement of the Parasympathetic Nervous System in Listening Effort. *Trends in Hearing*, 22, https://doi.org/10.1177/2331216518816603
- van den Brink, R. L., Murphy, P. R., & Nieuwenhuis, S. (2016). Pupil diameter tracks lapses of attention. *PLoS One*, 11(10), e0165274. doi:10.1371/journal.pone.0165274
- van der Leeuw, G., Leveille, S. G., Jones, R. N., Hausdorff, J. M., McLean, R., Kiely, D. K., ... Milberg, W. P. (2017). Measuring attention in very old adults using the Test of Everyday Attention. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, 24(5), 543–554. doi:10.1080/13825585.2016.1226747
- van der Wel P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, 25(6), 1–11. doi:10.3758/s13423-018-1432-y]
- Van Gerven, P. W. M., Paas, F., Merriënboer, J. J. G. V., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, 41(2), 167–174. doi:10.1111/j.1469-8986.2003.00148.x
- Wang, C.-A., Blohm, G., Huang, J., Boehnke, S. E., & Munoz, D. P. (2017). Multisensory integration in orienting behavior: Pupil size, microsaccades, and saccades. *Biological Psychology*, 129, 36–44. doi:10.1016/j.biopsycho.2017.07.024
- Wang, Y., Zekveld, A. A., Naylor, G., Ohlenforst, B., Jansma, E. P., Lorens, A., ... Kramer, S. E. (2016). Parasympathetic nervous system dysfunction, as identified by pupil light reflex, and its possible connection to hearing



- impairment. *PLoS One*, *11*(4), e0153566. doi:10.1371/journal.pone.0153566
- Wendt, D., Dau, T., & Hjortkjær, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Frontiers in Psychology*, *7*, 1–12. doi:10.3389/fpsyg.2016.00345
- Wendt, D., Hietkamp, R. K., & Lunner, T. (2017). Impact of noise and noise reduction on processing effort: A pupillometry study. *Ear and Hearing*, *38*(6), 690. doi:10.1097/AUD.0000000000000454
- Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8456–8460. doi:10.1073/pnas.1201858109
- Winn, B., Whitaker, D., Elliott, D. B., & Phillips, N. J. (1994). Factors affecting light-adapted pupil size in normal human subjects. *Investigative Ophthalmology & Visual Science*, *35*(3), 1132–1137.
- Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and Hearing*, *36*(4), e153–e165. doi:10.1097/AUD.0000000000000145
- Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in Hearing*, *22*. doi:10.1177/2331216518800869
- Woods, K. J. P., & McDermott, J. H. (2015). Attentive tracking of sound sources. *Current Biology*, *25*(17), 2238–2246. doi:10.1016/j.cub.2015.07.043
- Zaborszky, L., Csordas, A., Mosca, K., Kim, J., Gielow, M. R., Vadasz, C., & Nadasdy, Z. (2015). Neurons in the basal forebrain project to the cortex in a complex topographic organization that reflects corticocortical connectivity patterns: An experimental study based on retrograde tracing and 3D reconstruction. *Cerebral Cortex (New York, NY)*, *25*(1), 118–137. doi:10.1093/cercor/bht210
- Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in Hearing*, *22*, 1–25. doi:10.1177/2331216518777174
- 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*(4), 480. doi:10.1097/AUD.0b013e3181d4f251
- Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, *32*(4), 498. doi:10.1097/AUD.0b013e31820512bb