Behavioral/Cognitive

Eye Movements Decrease during Effortful Speech Listening

M. Eric Cui[1,2] **and Björn Herrmann**[1,2]

[1]Rotman Research Institute, Baycrest Academy for Research and Education, North York, Ontario M6A 2E1, Canada and [2]Department of Psychology, University of Toronto, Toronto, Ontario M5S 1A1, Canada

Hearing impairment affects many older adults but is often diagnosed decades after speech comprehension in noisy situations has become effortful. Accurate assessment of listening effort may thus help diagnose hearing impairment earlier. However, pupillometry—the most used approach to assess listening effort—has limitations that hinder its use in practice. The current study explores a novel way to assess listening effort through eye movements. Building on cognitive and neurophysiological work, we examine the hypothesis that eye movements decrease when speech listening becomes challenging. In three experiments with human participants from both sexes, we demonstrate, consistent with this hypothesis, that fixation duration increases and spatial gaze dispersion decreases with increasing speech masking. Eye movements decreased during effortful speech listening for different visual scenes (free viewing, object tracking) and speech materials (simple sentences, naturalistic stories). In contrast, pupillometry was less sensitive to speech masking during story listening, suggesting pupillometric measures may not be as effective for the assessments of listening effort in naturalistic speech-listening paradigms. Our results reveal a critical link between eye movements and cognitive load, suggesting that neural activity in the brain regions that support the regulation of eye movements, such as frontal eye field and superior colliculus, are modulated when listening is effortful.

Key words: **eye movements; eye-tracking; listening effort; pupillometry; speech processing; spoken stories**

Significance Statement

Assessment of listening effort is critical for early diagnosis of age-related hearing loss. Pupillometry is most used but has several disadvantages. The current study explores a novel way to assess listening effort through eye movements. We examine the hypothesis that eye movements decrease when speech listening becomes effortful. We demonstrate, consistent with this hypothesis, that fixation duration increases and gaze dispersion decreases with increasing speech masking. Eye movements decreased during effortful speech listening for different visual scenes (free viewing, object tracking) and speech materials (sentences, naturalistic stories). Our results reveal a critical link between eye movements and cognitive load, suggesting that neural activity in brain regions that support the regulation of eye movements are modulated when listening is effortful.

Introduction

Hearing impairment affects ~40% of people over 60 years of age (Feder et al., 2015; Goman and Lin, 2016) but is often diagnosed decades after speech comprehension difficulties in noisy situations, such as crowded restaurants, emerge (Pichora-Fuller et al., 1995; Pichora-Fuller and Levitt, 2012). Individuals with even mild hearing impairment rely substantially on attention in noisy situations, which makes listening effortful (Pichora-Fuller et al., 2016; Peelle, 2018; Herrmann and Johnsrude, 2020a). Accurate assessment of listening effort may thus help diagnose hearing impairment earlier and evaluate treatment outcomes. However, existing measures have limitations that hinder their use in practice. The current study aims to explore a novel approach to assess speech comprehension difficulties associated with listening effort.

Self-reports via subjective ratings are a common way to assess listening effort (Gatehouse and Noble, 2004; Larsby et al., 2005; Krueger et al., 2017). However, subjective measures can be influenced by different reference frames and attribution effects (Moore and Picou, 2018). Physiologic measures provide an objective window onto listening effort that can remedy disadvantages of subjective ratings (Mackersie et al., 2015; Wöstmann et al., 2015; Dimitrijevic et al., 2017; Miles et al., 2017). Pupillometry, the measurement of pupil dilation, is the most used objective tool to assess listening effort (Winn et al., 2018; Zekveld et al., 2018; Kadem et al., 2020; Neagu et al., 2023). However, pupil dilation is sensitive to environmental

Received Feb. 8, 2023; revised June 9, 2023; accepted July 18, 2023.

Author contributions: M.E.C. and B.H. designed research; M.E.C. and B.H. performed research; M.E.C. and B.H. analyzed data; and B.H. wrote the paper.

This work was supported by the Canada Research Chairs Grants CRC-2019-00156 and 232733, Natural Sciences and Engineering Research Council of Canada Grant RGPIN-2021-02602, Canadian Institutes of Health Research Grant 186236, and the William Demant Foundation. We thank Devin Sodums for help during data collection and Ingrid Johnsrude, Jennifer Ryan, and Lauren Fink for discussions.

The authors declare no competing financial interests.

Correspondence should be addressed to Björn Herrmann at bherrmann@research.baycrest.org.

https://doi.org/10.1523/JNEUROSCI.0240-23.2023

changes in luminance (Knapen et al., 2016) and light spectrum (Suzuki et al., 2019; Thurman et al., 2021) and is therefore difficult to use outside the laboratory even with portable eye-tracking equipment. Moreover, measuring pupil dilation accurately requires participants to continuously fixate on a point on a computer monitor (Ohlenforst et al., 2017; Zekveld et al., 2018; Farahani et al., 2020; Winn and Teece, 2021) because luminance changes arising from eye movements change the pupil dilation, and different angles of the pupil relative to the eye tracker can make the pupil diameter appear different without an actual difference (Brisson et al., 2013; Hayes and Petrov, 2016; Fink et al., 2023). Yet, restriction of gaze to a central fixation point creates an additional task requirement that is uncommon in everyday life, and central fixation can impair memory and mental imagery for spoken speech (Johansson et al., 2012).

A few attempts have been made to assess listening effort using the small jerk-like, involuntary eye movements during fixation (microsaccades) that can be measured concurrently with pupillometry (Kadem et al., 2020). However, microsaccades appear sensitive neither to acoustically nor linguistically induced listening efforts (Kadem et al., 2020; although Contadini-Wright et al., 2023, report microsaccade sensitivity to listening effort). Microsaccade amplitudes are also very small (Martinez-Conde et al., 2009, 2013), requiring high-resolution recording (Poletti and Rucci, 2016; Zhao et al., 2019a), and microsaccade recordings suffer from the same disadvantages associated with forced gaze fixation as does pupillometry. These challenges hinder the utility of pupillometric and microsaccadic indices for the assessment of speech comprehension difficulties in noise and associated listening effort.

Individuals naturally explore their environments through eye movements (Fukushima, 2003; Ono, 2015; Missal and Heinen, 2017), and visual exploration can be incidental while engaged in a different task (Lipton et al., 1980; Hutton and Tegally, 2005; Kosch et al., 2018). Such incidental eye movements may provide a window onto cognition as people avert gaze (Glenberg et al., 1998), reduce object-tracking eye movements (Lipton et al., 1980; Hutton and Tegally, 2005; Kosch et al., 2018), and decrease saccades (Walter and Bex, 2021) during periods of high memory load, raising the possibility that all cognitively demanding tasks, including speech comprehension, affect eye movements. This possibility is supported by neurophysiological evidence showing that a reduction in movements increases neural activity in the auditory cortex and, in turn, improves sound perception (Schneider et al., 2014; McGinley et al., 2015; Schneider and Mooney, 2015, 2018). Critically, eye movements directly modulate neuronal excitability in auditory cortex such that the likelihood of a neuron firing, and thus responding to sound, increases in the absence of eye movements (O'Connell et al., 2020). Research further suggests that eye movements and pupil dilation might be driven by common underlying neurophysiology (Joshi and Gold, 2020; Wang and Munoz, 2021; Burlingham et al., 2022) and, as a result, both may perhaps be sensitive to listening effort.

In the current study, we propose that leveraging eye movements to make inferences about audition will deliver an effective measure of listening effort. We suggest that when listening becomes effortful, eye movements decrease to free resources for speech comprehension. In three experiments with different speech materials and different visual-stimulation displays, we examine the hypothesis that eye movements are sensitive to the speech masking that is associated with listening effort.

Materials and Methods

Participants

Younger adults, age 18–34 years, participated in the three experiments of the current study. Participants were either native English speakers or highly proficient non-native English speakers. Demographic information for each participant is provided in the sections describing the methods for each experiment. Participants gave written informed consent before the experiment and were paid \$7.5 Canadian per half-hour for their participation. Participants self-reported having normal hearing abilities. The study was conducted in accordance with the Declaration of Helsinki and the Canadian Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS2-2014) and was approved by the Research Ethics Board of the Rotman Research Institute.

Experimental setup

Sounds were presented via Sony Dynamic Stereo MDR-7506 headphones and a Steinberg UR22mkII (Steinberg Media Technologies) external sound card. Experimental procedures were run using Psychtoolbox software (version 3.0.14) in MATLAB (MathWorks) on a Lenovo T450s laptop with Microsoft Windows XP software. The laptop screen was mirrored to an ASUS monitor with a refresh rate of 60 Hz. All sounds were presented at a comfortable listening level that was fixed across participants (~70–75 dB SPL).

During the experiments, participants rested their head on a chin and forehead rest facing the computer monitor at a distance of ~70 cm. Pupil area and eye movements were recorded continuously from the right eye (or the left eye if the right eye could not be tracked accurately) using an integrated infrared camera (EyeLink 100 Plus eye tracker, SR Research) at a sampling rate of 500 Hz. Nine-point fixation was used for eye-tracker calibration before each block (McIntire et al., 2014).

Preprocessing of eye-movement and pupil-area data

Preprocessing of eye-movement and pupil-area data involved removing eye blinks and other artifacts. For each eye blink indicated by the eye tracker, all data points between 100 ms before and 200 ms after a blink were set to NaN ("not a number" in MATLAB). In addition, pupil-area values that differed from the mean pupil area by >3 times the SD were classified as outliers and set to NaN. Missing pupil data (coded as NaN) resulting from artifact rejections and outlier removal were interpolated using the MATLAB pchip method; x-time and y-time courses were not interpolated (except for microsaccade/saccade analyses), but missing data points (NaNs) were ignored in eye-movement data analysis procedures.

Pupil- and eye-movement metrics

Pupil-area time courses were filtered with a 5 Hz low-pass filter (51 points, Kaiser window). Because participants were not required to fixate in one location on the screen for most of the experimental conditions, pupil area could be affected by the changing angle of the pupil relative to the eye-tracking camera. To mitigate this potential issue, we regressed out any linear and quadratic relationship of x and y with the pupil area before low-pass filtering (Fink et al., 2023; Kinley and Levy, 2022; Kraus et al., 2022). That is, for the continuous data of each block, we calculated a regression using the pupil-area time course as the dependent variable and the time courses of the x, y, squared x, and squared y coordinates as predictors. The residual pupil area was used as the pupil-area measure in the current study in MATLAB code as follows: `[~,~,pres] = regress(p, [x y x.^2 y.^2 ones(length(p), 1)]`, where p is the pupil-area time course, and x and y the horizontal and vertical eye-tracking time courses, respectively; pres is the residual pupil area. Critically, there were no meaningful differences in the results reported below for the corrected pupil area compared with the uncorrected pupil area (that is, the same effects/interactions were statistically significant for both).

Two main metrics were used to investigate whether eye movements changed depending on the degree of speech masking—fixation duration and spatial gaze dispersion (MATLAB code, https://osf.io/pb4dr/). Fixation duration was calculated as the time a person's x–y eye coordinates remained in a given location (within 0.5° visual angle, radius of 10 pixels). For each time point, the corresponding x–y coordinate defined the critical 0.5° visual-angle location. The number of continuous presamples and postsamples was calculated for the x–y eye coordinates that remained in the critical location. The sample number was divided by the sampling frequency to obtain the fixation duration for the specific time point. If a data value of any presample or postsample within the 0.5° visual angle location had been coded NaN (i.e., was missing), the fixation duration of the corresponding time point was set to NaN and ignored during averaging.

Spatial gaze dispersion is a measure of the general tendency for the eyes to move around. It was calculated as the SD in gaze across time points, averaged across x and y coordinates, and transformed to logarithmic values. Smaller values indicate less gaze dispersion. To obtain time courses for gaze dispersion, it was calculated for 1 s sliding time windows centered sequentially on each time point. If >90% of data were unavailable within a 1 s time window (i.e., >450 samples were NaN coded), gaze dispersion for the corresponding time point was set to NaN and ignored during averaging.

Fixation duration and gaze dispersion do not make any assumptions about the type of eye movements under investigation and may thus be uniquely sensitive to individuals listening to masked speech. Nevertheless, we also analyzed the saccade/microsaccade rate using a method that computes thresholds based on velocity statistics from x and y coordinate trial time courses (NaN-coded data were interpolated) and then identifies saccades/microsaccades as events passing that threshold (Engbert and Kliegl, 2003; Engbert, 2006; Widmann et al., 2014; Zhao et al., 2019a; Contadini-Wright et al., 2023). That is, the vertical and horizontal eye movement time series were transformed into velocities, separately for each trial, and saccades/mircosaccades were classified as outliers if they exceeded a relative velocity threshold of several times the median SD of the eye-movement velocity and persisted for 6 ms or longer. Previous work differed in the specific threshold that was used. Some works used a velocity threshold of five times the median SD of the eye movement velocity (Engbert and Kliegl, 2003; Widmann et al., 2014), whereas others used a threshold of 15 times the median SD (Kadem et al., 2020). A lower threshold leads to a higher number of data points that are considered a saccade/microsaccade. It is unclear whether the threshold may affect the sensitivity to speech masking. Hence, we calculated analyses for both thresholds. A time course of saccade/microsaccade rate was calculated from the individual saccade/microsaccade times (Widmann et al., 2014; Zhao et al., 2019a; Kadem et al., 2020) by convolving each occurrence with a Gaussian function (SD of 0.02 s; zero phase lag). We do not distinguish between saccades and microsaccades because the definition depends on the velocity threshold, and previous work suggests that the same mechanisms underlie saccades and microsaccades (Martinez-Conde et al., 2009, 2013).

Statistical analysis

The experiments reported here were not preregistered. Experimental manipulations were within-participants factors. Differences between experimental conditions were thus assessed using one-sample t tests, paired-samples t tests, and repeated-measures ANOVAs (rmANOVAs). Reporting of statistical results includes test statistic, degrees of freedom, significance level, and effect size. Note that nonparametric statistics yielded qualitatively similar results. Details about statistical analyses are provided for each experiment separately below. Effect sizes for rmANOVAs and t tests are reported as omega squared (ω^2) and Cohen's d (d), respectively. All statistical analyses described were conducted using MATLAB (MathWorks) and JASP (version 0.16.34) software.

Experiment 1: the influence of speech masking on eye movements

Central fixation is common in pupillometry studies (Kuchinsky et al., 2013; Ohlenforst et al., 2017; Zekveld et al., 2018; Farahani et al., 2020; Kadem et al., 2020; Winn and Teece, 2021), but the restricted gaze may hinder examining the sensitivity of eye movements to listening effort. Moreover, forced fixation reflects an additional task that may draw resources from speech comprehension. Experiment 1 explores the impact of fixation relative to free viewing on speech comprehension and the degree to which eye movements index speech comprehension difficulties.

Participants

Twenty-six adults (median age, 23.5 years; age range, 18–30 years; 19 female, 7 male) participated in experiment 1. Data from two additional participants were excluded because not all blocks were recorded because of technical issues. For four of the 26 participants, the quality of the eye-tracking and pupil data were low. That is, >30% of trials contained >40% of missing data (Kadem et al., 2020). Hence, data from 22 adults (median age, 23 years; age range, 18–30 years; 16 female, 6 male) were available for analyses of eye movements ($N = 26$ for behavioral analysis). Seven of the 22 participants were native English speakers, the other 15 participants were highly proficient non-native English speakers. [The mean self-rated English skills of non-native English speakers on a 0 (poor) to 10 (very high) scale was 8.9.] Several of the participants who indicated having a non-English first language grew up in English-speaking Canada and have been speaking English since early childhood (<5 years of age). Because our manipulations were all within-participants factors, the participants' language status does not confound our speech-masking investigation.

Stimulus materials and procedure

Participants listened to short sentences spoken by a female native English speaker (mean duration, 2.5 s; range of durations, 2–3.2 s). Sentences were embedded in a 7 s 12-talker babble noise (Bilger, 1984). A sentence started 3 s after babble onset (Zhang et al., 2022). Sentences were presented either at −2 dB, +3 dB, or +8 dB signal-to-noise ratio (SNR). Different SNRs were achieved by adjusting the sound level of the sentence while keeping the babble level constant across trials (Ohlenforst et al., 2017; Kadem et al., 2020). This ensured that stimuli with different SNRs did not differ before sentence onset (Ohlenforst et al., 2017; Kadem et al., 2020). Assignment of SNR levels to specific sentences was randomized across participants. After each stimulus, a probe word appeared on the screen. Participants were asked to indicate whether the probe word is semantically related or unrelated to the sentence (Rodd et al., 2010a, b; Kadem et al., 2020). At the same time as participants were hearing the auditory stimuli, they were either presented with a blank gray screen (free viewing) or with a fixation square centered on the gray screen (fixation; the onset of the fixation square coincided with the onset of the babble and stayed on the screen until babble offset).

Participants were presented with six experimental blocks. In three blocks, participants were presented with the blank screen (free viewing), whereas in the other three blocks, participants were presented with the fixation square. Blocks with the blank screen versus fixation square alternated, and the starting block type was counterbalanced across participants. In each block, participants listened to stimuli in eight trials per SNR level, presented in pseudorandomized order, such that a maximum of three trials of the same SNR level could occur in a row. Hence, across the experiment, participants listened to 24 trials for each SNR level and each fixation condition.

To examine whether participants found free viewing less exhausting than fixating across a block, participants rated their mental workload for six statements after each block using the following comments: "I felt this block went by slowly," "I found this block exhausting," "I feel tired of listening," "I had to invest a lot of effort during this block," "I felt the block was mentally demanding," and "I look forward to a short break." Participants rated each of these statements using a 7-point scale, where 1 referred to strongly disagree and 7 referred to strongly agree. No difference in the averaged ratings was found between fixation and free viewing ($t_{(25)} = 0.977$, $p = 0.338$, $d = 0.192$).

Analysis of behavioral data

The proportion of correct responses in the semantic-relatedness task was calculated for each SNR condition. An rmANOVA was calculated using the proportion of correct responses as a dependent measure. Within-participants factors were SNR (−2, +3, +8 dB SNR) and Viewing Condition (free viewing, fixation).

Analysis of eye-movement and pupil-area data

Continuous pupil-area, x coordinate and y coordinate data were divided into single-trial time courses ranging from −1 to 7 s time locked to babble onset. Data for an entire trial were excluded from analysis if the percentage of NaN data entries made up >40% of the trial (Kadem et al., 2020).

Fixation duration, gaze dispersion, saccade/microsaccades rate, and pupil area were calculated for each trial and then averaged across trials separately for each SNR condition. For statistical analyses, fixation duration, gaze dispersion, and saccade/microsaccades rate were averaged across the 3–5.5 s time window during which sentences were presented (sentences started at 3 s after babble onset; average sentence offset was at 5.5 s). Mean pupil area was calculated for the 4–6.5 s (i.e., delayed by 1 s), because the pupil dilation is known to change relatively slowly and peak late during sentence listening (Knapen et al., 2016; Winn and Moore, 2018; Winn et al., 2018; Kadem et al., 2020; Zhang et al., 2022). An rmANOVA with the within-participants factors SNR (−2, +3, +8 dB SNR) and Viewing Condition (free viewing, fixation) was calculated separately for fixation duration, gaze dispersion, saccade/microsaccade rate, and pupil area.

Results

Speech comprehension differs between free viewing and fixation

Behavioral performance in the semantic-relatedness task decreased with decreasing SNR ($F_{(2,50)} = 57.918$, $p = 9.6 \cdot 10^{-14}$, $\omega^2 = 0.511$; no effect of Viewing Condition, $F_{(2,50)} = 0.003$, $p = 0.957$, $\omega^2 < 0.001$), but this SNR-related reduction differed depending on whether individuals freely moved or fixated their eyes (SNR × Viewing Condition interaction, $F_{(2,50)} = 4.903$, $p = 0.011$, $\omega^2 = 0.062$; Fig. 1). The reduction in performance from +8 dB to −2 dB SNR was greater for free viewing than fixation ($t_{(25)} = 2.886$, $p = 0.008$, $d = 0.566$), showing that speech comprehension scores discriminate less between speech-masking conditions during fixation than free viewing. Interestingly, under the most difficult listening condition (−2 dB SNR), a reduction in eye movements (fixation) was associated with better speech comprehension relative to free viewing ($t_{(25)} = 2.186$, $p = 0.038$, $d = 0.429$), which is consistent with the hypothesis that reduced eye movements support listening under challenges.

Eye movements decrease and pupil area increases with increasing speech masking

Fixation durations were shorter during free viewing than during fixation, as expected ($F_{(1,21)} = 12.069$, $p = 0.002$, $\omega^2 = 0.049$). Critically, fixation durations increased with decreasing SNR ($F_{(2,42)} = 5.752$, $p = 0.006$, $\omega^2 = 0.006$; Fig. 2A). Specifically, fixation durations were longer for −2 dB SNR compared with +3 dB SNR ($t_{(21)} = 2.201$, $p = 0.039$, $d = 0.469$) and +8 dB SNR ($t_{(21)} = 3.423$, $p = 0.003$, $d = 0.730$), whereas fixation durations did not differ between +3 and +8 dB SNR ($t_{(21)} = 0.884$, $p = 0.387$, $d = 0.188$). There was no SNR × Viewing Condition interaction ($F_{(2,42)} = 0.056$, $p = 0.946$, $\omega^2 < 0.001$).

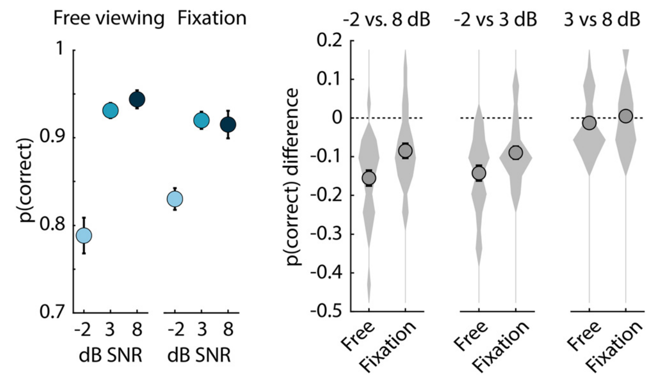

Figure 1. Behavioral results for experiment 1. Proportion of correct responses in the semantic-relatedness task (left) and the difference in the proportion of correct responses between different SNRs (right). The reduction in speech comprehension for −2 dB relative to 8 dB SNR was greater for free viewing than fixation ($p < 0.05$). Speech comprehension was also greater for fixation than free viewing under challenging listening conditions (−2 dB SNR; $p < 0.05$). Violin plots reflect the histogram of individual data points. Error bars indicate SEM.

The results for gaze dispersion were consistent with those for fixation duration. Gaze dispersion was lower during fixation than free viewing as expected ($F_{(1,21)} = 53.229$, $p = 3.5 \cdot 10^{-7}$, $\omega^2 = 0.213$). Gaze dispersion decreased with decreasing SNR, although this effect was only marginally significant ($F_{(2,42)} = 3.208$, $p = 0.051$, $\omega^2 = 0.001$; Fig. 2B). Gaze dispersion was smaller for −2 dB SNR compared with +8 dB SNR ($t_{(21)} = 2.620$, $p = 0.016$, $d = 0.559$), whereas the other contrasts were not significant (p values > 0.16). The SNR × Viewing Condition interaction was marginally significant ($F_{(2,42)} = 2.841$, $p = 0.070$, $\omega^2 < 0.001$) because gaze dispersion was only affected by SNR during free viewing ($F_{(2,42)} = 4.050$, $p = 0.025$, $\omega^2 = 0.003$) but not during fixation ($F_{(2,42)} = 0.684$, $p = 0.510$, $\omega^2 < 0.001$).

As expected based on previous work (Zekveld et al., 2010; Zekveld and Kramer, 2014; Winn et al., 2015; Winn, 2016; Winn et al., 2018; Zekveld et al., 2018; Kadem et al., 2020), pupil area increased with increasing speech masking (main effect of SNR, $F_{(2,42)} = 11.680$, $p = 9.3 \cdot 10^{-5}$, $\omega^2 = 0.035$). The pupil area was larger for −2 dB SNR compared with +3 dB SNR ($t_{(21)} = 3.034$, $p = 0.006$, $d = 0.647$) and +8 dB SNR ($t_{(21)} = 4.487$, $p = 2 \cdot 10^{-4}$, $d = 0.957$). Pupil area did not differ between +3 and +8 dB SNR ($t_{(21)} = 1.806$, $p = 0.085$, $d = 0.385$). Pupil area was also smaller during fixation than free viewing (main effect of Viewing Condition, $F_{(1,21)} = 8.967$, $p = 0.007$, $\omega^2 = 0.037$). The SNR × Viewing Condition interaction was not significant ($F_{(2,42)} = 1.530$, $p = 0.228$, $\omega^2 = 0.002$).

We also investigated whether the reduction in eye movements is specifically related to saccadic/microsaccadic eye movements (Fig. 3). Saccade/microsaccade rate decreased with increasing speech masking but only for the high- and not the low-velocity threshold defining a saccade/microsaccade. Specifically, the rmANOVA for the saccade/microsaccade rate calculated using a threshold of five times the median

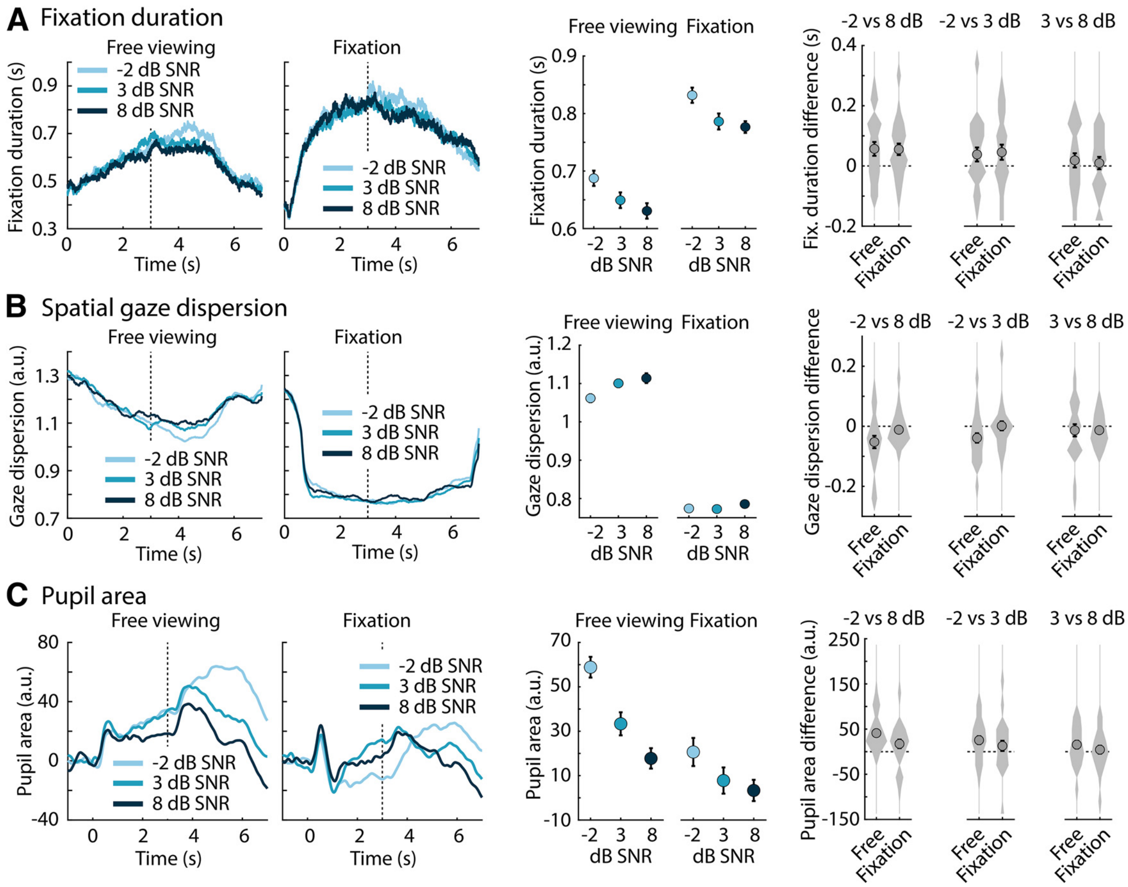

Figure 2. Fixation duration, spatial gaze dispersion, and pupil area are sensitive to speech masking. ***A***, Left, Fixation duration time courses while participants freely viewed a blank computer screen or fixated on a fixation square at the center of the screen. The vertical dashed line in the time course plots indicates the sentence onset. Middle, Mean fixation duration (across the 3–5.5 s time window; sentence onset at 3 s and average sentence offset at 5.5 s). Error bars indicate SEM (removal of between-participant variance; Masson and Loftus, 2003). Right, Difference in fixation duration between different SNR levels for each viewing condition (free viewing, fixation). Violin plots reflect the histogram of individual data points. Dots reflect the mean. ***B***, Same as in ***A*** for spatial gaze dispersion. ***C***, Same as in ***A*** for pupil area (mean across the 4–6.5 s time window).

SD revealed no effect of Viewing Condition ($F_{(1,21)} = 0.012$, $p = 0.913$, $\omega^2 < 0.001$) nor of SNR ($F_{(2,42)} = 1.098$, $p = 0.343$, $\omega^2 < 0.001$) nor a SNR × Viewing Condition interaction ($F_{(2,42)} = 2.186$, $p = 0.125$, $\omega^2 = 0.002$). The rmANOVA for the saccade/microsaccade rate calculated using a threshold of 15 times the median SD revealed a main effect of SNR ($F_{(2,42)} = 4.617$, $p = 0.015$, $\omega^2 = 0.007$), whereas the effect of Viewing Condition ($F_{(1,21)} = 0.038$, $p = 0.847$, $\omega^2 < 0.001$) and the SNR × Viewing Condition interaction ($F_{(2,42)} = 0.812$, $p = 0.451$, $\omega^2 < 0.001$) were not significant. The saccade/microsaccade rate was lower for the −2 dB compared with the +8 dB condition ($t_{(21)} = 3.162$, $p = 0.005$, $d = 0.674$). The −2 dB versus +3 dB SNR contrast was marginally significant ($t_{(21)} = 1.896$, $p = 0.072$, $d = 0.404$), whereas no difference was found between +3 dB and +8 dB ($t_{(21)} = 0.849$, $p = 0.406$, $d = 0.181$).

We further investigated whether SNR-related changes in eye movements and changes in pupil area may be driven by the same underlying mechanism. To this end, we subtracted the +8 dB condition from the −2 dB condition [averaged across viewing conditions (free, fixation)], separately for pupil area, fixation duration, and spatial gaze dispersion. We calculated Pearson correlations between the difference in pupil area and the difference in fixation duration and gaze dispersion, but no significant correlation was found (fixation duration, $r = 0.103$, $p = 0.647$; gaze dispersion, $r = -0.256$, $p = 0.250$), perhaps suggesting somewhat independent processes.

Finally, the SNR effect (−2 dB minus +8 dB SNR) for behavioral performance did not correlate with the SNR effect for pupil area, fixation duration, nor gaze dispersion (fixation duration, $r = -0.005$, $p = 0.822$; gaze dispersion, $r = -0.202$, $p = 0.368$; pupil area, $r = 0.377$, $p = 0.084$), suggesting effort measures and behavioral performance dissociate to some extent, which is in line with Koelewijn et al. (2018) and Carolan et al. (2022).

Summary

Experiment 1 shows that eye movements decrease when listening is effortful because of speech masking (Fig. 2) and that speech comprehension under challenging listening is better when individuals reduce their eye movements (Fig. 1). Our data also show that speech comprehension scores discriminate less between speech masking conditions during fixation than free viewing. (Fig. 1). The results thus suggest that eye movements provide a window onto the perceptual/cognitive load during masked speech listening. The current results further suggest that eye movements could potentially be used to assess listening effort under nonfixation conditions and conditions that involve visual exploration. We conducted experiment 2 to test directly whether eye movements are also sensitive to speech masking when individuals can engage in visual exploration.

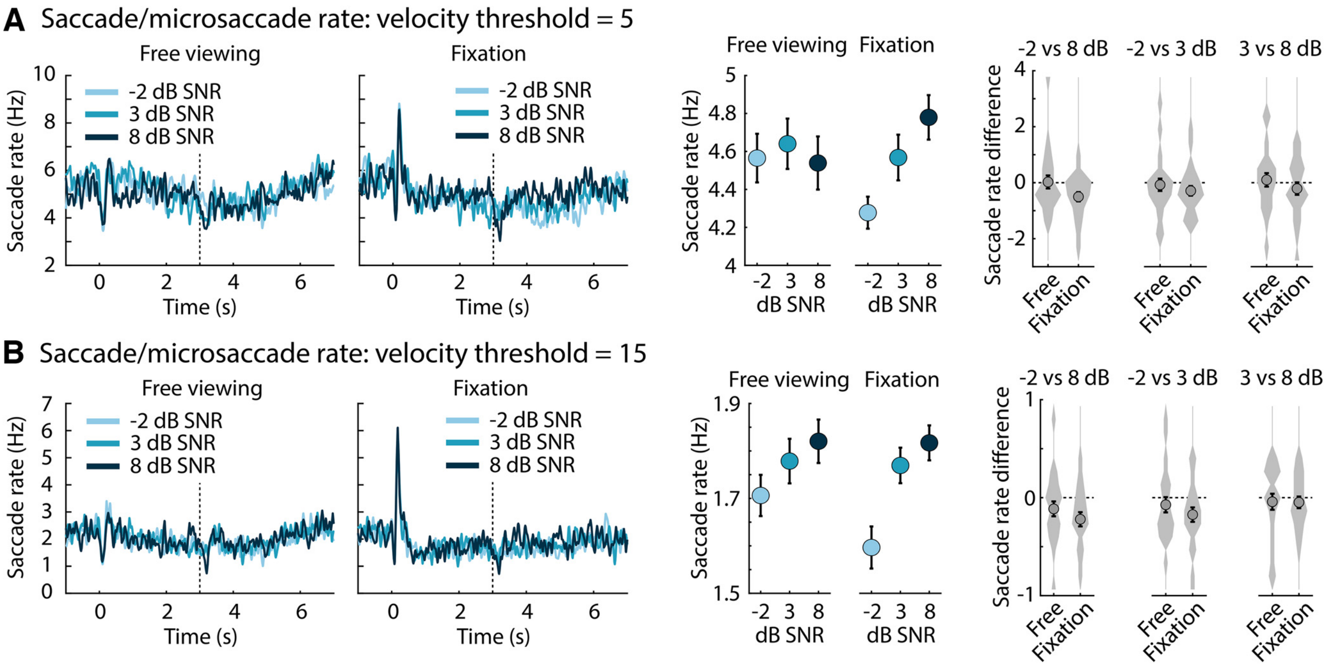

Figure 3. Results for saccade/microsaccade rate. ***A***, Left, Saccade/microsaccade rate time courses during free viewing or fixation. Saccade/microsaccade rate calculated using a velocity threshold of 5 times the median SD of the eye-movement velocity. The vertical dashed line in the time course plots indicates the sentence onset. Middle, Mean saccade/microsaccade rate (across the 3–5.5 s time window). Error bars indicate SEM (removal of between-participant variance; Masson and Loftus, 2003). Right, Difference in saccade/microsaccade rate between different SNR levels for each viewing condition (free viewing, fixation). Violin plots reflect the histogram of individual data points. Dots reflect the mean across participants. ***B***, Same as in ***A*** for saccade/microsaccade rate calculated using a velocity threshold of 15 times the median SD of the eye-movement velocity.

Experiment 2: the Influence of speech masking on eye movements during incidental object tracking

Participants

Twenty-two adults (median age, 23 years; age range, 18–32 years; 14 female, 7 male; 1 person did not provide demographic information but was recruited from the same participant pool) participated in experiment 2. Data from one additional participant were recorded but excluded from analysis because the person's behavioral performance was at chance level even for the easy speech comprehension condition. Fifteen of the 22 participants were native English speakers, the other 6 participants were highly proficient non-native English speakers (the one person who did not provide demographic information was a highly proficient, likely native, English speaker). The mean self-rated English skills of non-native English speakers on a 0 (poor) to 10 (very high) scale was 8.3.

Stimulus materials and procedure

Recordings of sentences from the Harvard sentence lists (mean duration, 2.4 s; range of durations, 1.9–2.9 s; Institute of Electrical and Electronics Engineers, 1969) spoken by a male native English speaker were used in experiment 2. Sentences were embedded in a 6 s 12-talker babble noise (Bilger, 1984). A sentence started 2.5 s after babble onset. Sentences were presented either at −3 dB SNR or at +10 dB SNR. As for experiment 1, different SNRs were achieved by adjusting the sound level of the sentence while keeping the babble level constant across trials (Ohlenforst et al., 2017; Kadem et al., 2020). Assignment of SNR levels to specific sentences was randomized across participants. After each stimulus, a probe word occurred on the screen, and participants performed the semantic-relatedness judgment task (Rodd et al., 2010a, b; Kadem et al., 2020).

At the same time as participants were hearing the auditory stimuli, they were presented with an incidental object-viewing display (see Fig. 4*A*). Before babble onset, a dot was presented at the center of the screen for 0.5 s at a suprathreshold background-to-dot contrast. The dot during this 0.5 s period was colored yellow or green depending on the sentence-to-babble SNR. That is, the color of the dot served as a cue to uniquely indicate the sentence-comprehension difficulty (100% valid cue). The assignment of the colors yellow and green to the SNR levels −3 dB and +10 dB was counterbalanced across participants. Many participants reported not using the color cue; the color differences may have been too subtle. On babble onset, the dot turned to a light gray and started to move in a random, smooth trajectory for the 6 s duration of a trial (within a 14° visual angle; see Fig. 4*A*). Participants were instructed that the task was to comprehend the sentence so that they would be able to decide whether the probe word was semantically related or unrelated to the sentence. No task was associated with the object-movement display. Participants were instead instructed to look at the screen in whatever way they wanted (Johansson et al., 2006, 2011, 2012).

Participants were presented with four blocks. In each block, participants listened to 18 trials of each of the two SNR conditions presented in pseudorandomized order such that a maximum of three trials of the same SNR condition were presented in a row. Hence, across the experiment, participants listened to 72 trials per SNR condition (18 trials × 4 blocks).

Analysis of behavioral data

The proportion of correct responses in the semantic-relatedness task was calculated for each SNR condition. A dependent-samples *t* test was used to compare the proportion of correct responses between the −3 dB and the +10 dB SNR condition.

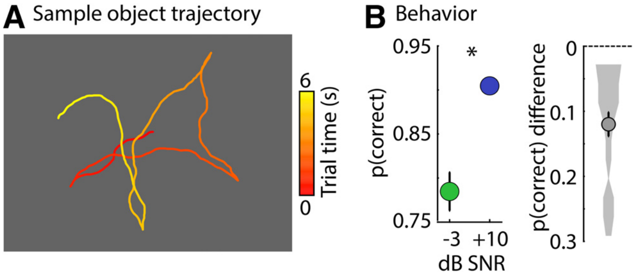

Figure 4. Visual stimulation and behavioral results for experiment 2. ***A***, Sample of an object-movement trajectory over the 6 s duration of a trial. Color indicates the temporal evolution of the movement trajectory. ***B***, Left, Proportion of correct responses in the semantic-relatedness task. Error bars indicate SEM. Right, Mean difference in the proportion of correct responses between the two SNR levels (−3 dB SNR minus +10 dB SNR). The violin plot reflects the histogram of individual data points; $^{*}p \leq 0.05$.

Analysis of eye-movement and pupil-area data

Continuous pupil-area, x coordinate, and y coordinate data were divided into single-trial time courses ranging from −1-6 s time locked to babble onset. Data for an entire trial were excluded from analysis if the percentage of NaN data entries made up >40% of the trial (Kadem et al., 2020).

Fixation duration was calculated as described above. Spatial gaze dispersion was calculated slightly differently in experiment 2 than for experiment 1 to account for variance associated with the movement of the object. Specifically, the SD in gaze across time points (averaged across x and y coordinates) and the SD in object movement across time points (averaged across x and y coordinates) were calculated separately. The resulting SD of the gaze was then divided by the SD of the object movement, and the result was transformed to logarithmic values. Note that the results for the normalized gaze dispersion were qualitatively similar to the results for non-normalized gaze dispersion (calculated as for experiment 1). In addition to fixation duration and gaze dispersion, we also calculated Pearson's correlation between the x and y coordinates of the object and the eye, and transformed the mean result to Fisher's z scores. To obtain time courses of the eye metrics, fixation duration was calculated for each time point, and gaze dispersion and object tracking were calculated for 1 s sliding time windows centered sequentially on each time point.

Time courses for fixation duration, gaze dispersion, Fisher's z scores (object tracking), saccade/microsaccade rate, and pupil area were averaged across trials. For statistical analyses, fixation duration, gaze dispersion, Fisher's z scores, and saccade/microsaccade rate were averaged across the 2.5–4.9 s time window during which sentences were presented (sentences started at 2.5 s after babble onset; average sentence offset was at 4.9 s). For the analysis of the pupil area, the mean pupil area in the 3.5–5.9 s time window was calculated (again shifted by 1 s to account for delayed responsivity of the pupil; Knapen et al., 2016; Winn and Moore, 2018; Winn et al., 2018; Zhang et al., 2022). A dependent-samples t test was calculated to compare fixation duration, gaze dispersion, Fisher's z scores, saccade/microsaccade rate, and pupil area between the two SNR levels (−3, +10 dB SNR).

Results

Behavioral data are displayed in Figure 4*B*. The proportion of correct responses in the semantic-relatedness task was lower for the −3 dB SNR compared with the +10 dB SNR condition ($t_{(21)} = 6.494$, $p = 2 \cdot 10^{-6}$, $d = 1.385$), as expected.

Analyses for eye movements showed that fixation duration increased ($t_{(21)} = 2.775$, $p = 0.011$, $d = 0.592$), and both gaze dispersion ($t_{(21)} = 2.824$, $p = 0.010$, $d = 0.602$) and object tracking decreased ($t_{(21)} = 2.798$, $p = 0.011$, $d = 0.597$) for the more challenging (−3 dB SNR) compared with the more favorable SNR (+10 dB SNR; Fig. 5*A–C*). These data show that eye movements decrease when speech masking increases and makes listening effortful during incidental object tracking. Pupil area was larger for speech presented at −3 dB compared with +10 dB SNR, but this was only marginally significant ($t_{(21)} = 1.794$, $p = 0.087$, $d = 0.382$; Fig. 5*D*).

The saccade/microsaccade rate was lower for −3 dB compared with +10 dB SNR, but only for the low velocity threshold that defines a saccade/microsaccade ($t_{(21)} = 2.374$, $p = 0.027$, $d = 0.506$; Fig. 6*A*) and not for the high-velocity threshold ($t_{(21)} = 1.497$, $p = 0.149$, $d = 0.319$; Fig. 6*B*). This contrasts with experiment 1, where the SNR effect was only significant for the high-velocity threshold.

Similar to observations in experiment 1, SNR-related changes (−3 dB minus +10 dB SNR) in pupil area did not correlate with changes in eye-movement metrics (fixation duration, $r = 0.255$, $p = 0.253$; gaze dispersion, $r = -0.287$, $p = 0.253$; object tracking, $r = -0.374$, $p = 0.087$), nor did the SNR effect in behavioral performance correlate with eye-movement metrics or pupil area (fixation duration, $r = -0.066$, $p = 0.771$; gaze dispersion, $r = 0.02$, $p = 0.93$; object tracking, $r = 0.005$, $p = 0.984$; pupil area, $r = 0.182$, $p = 0.418$).

Summary

Experiment 2 expands the results of experiment 1, showing that eye movements decrease when speech masking increases under incidental object tracking. The results of experiments 1 and 2 provide a clear demonstration that eye movements carry critical information about the challenges during speech comprehension and that this sensitivity of eye movements to masked speech listening generalizes to different visual-stimulation conditions. Importantly, listening situations not only vary in the visual information available but also in the type of speech materials. Experiment 3 examines whether eye movements are also sensitive to listening effort for engaging, continuous speech materials that are common in everyday life (Jefferson, 1978; Mullen and Yi, 1995; Bohanek et al., 2009; Fivush et al., 2011).

Experiment 3: the influence of speech masking on eye movements during story listening

Difficulties with speech comprehension and associated effort have been assessed mostly using brief disconnected sentences (Zekveld et al., 2010; Wendt et al., 2016; Ayasse and Wingfield, 2018; Zekveld et al., 2019; Kadem et al., 2020; Winn and Teece, 2021; compare experiments 1 and 2). Such materials lack a topical thread and are not very interesting to a listener. Speech in everyday life is often continuous and follows an overarching theme, and a listener is intrinsically motivated to comprehend (Jefferson, 1978; Mullen and Yi, 1995; Bohanek et al., 2009; Fivush et al., 2011; Herrmann and Johnsrude, 2020a). Any measure of listening effort would ideally be sensitive to such continuous, naturalistic speech. However, pupillometry may be less suited for the assessment of listening effort during continuous speech listening because measures of pupil dilation typically require normalization to a presentence baseline period (Zekveld et al., 2010; Winn et al., 2018; Zekveld et al., 2018, 2019; Kadem et al., 2020; Winn and Teece, 2021), which is less possible for continuous speech (but see studies that used ∼30-s stimuli, Zhao et al., 2019b; Fiedler et al.,

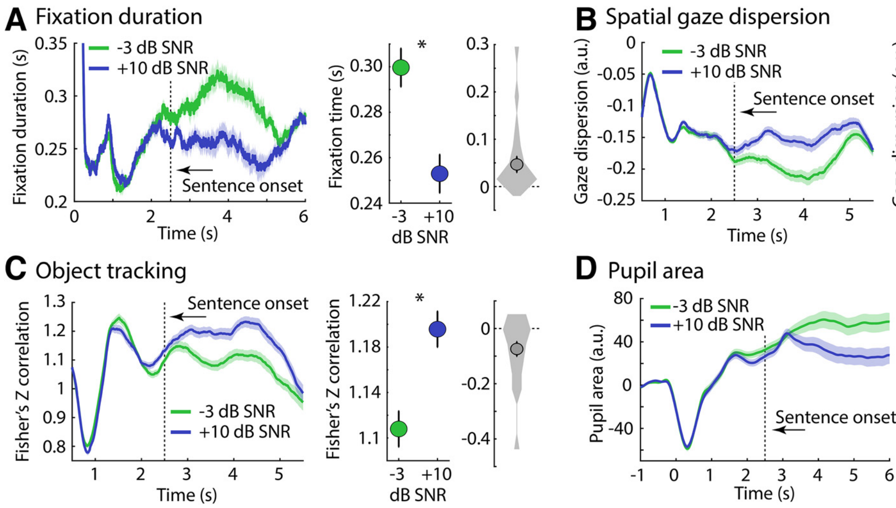

Figure 5. Fixation duration, gaze dispersion, object tracking, and pupil area are sensitive to speech masking. ***A***, Left, Fixation duration time courses. The vertical dashed line marks the sentence onset. Middle, Mean fixation duration (across the 2.5–4.9 s time window; sentence onset at 2.5 s and average sentence offset at 4.9 s). Error bars and shading indicate SEM (removal of between-participant variance; Masson and Loftus, 2003). Right, Difference in fixation duration between the two SNR levels (−3 dB SNR minus +10 dB SNR). The violin plot reflects the histogram of individual data points. The dot reflects the mean across participants. ***B***, Same as in ***A*** for spatial gaze dispersion. ***C***, Same as in ***A*** for Fisher's *z*-transformed correlation between object and eye position (object tracking). ***D***, Same as in ***A*** for pupil area (mean across the 3.5–5.9 s time window); $*p \leq 0.05$, $\#p \leq 0.1$.

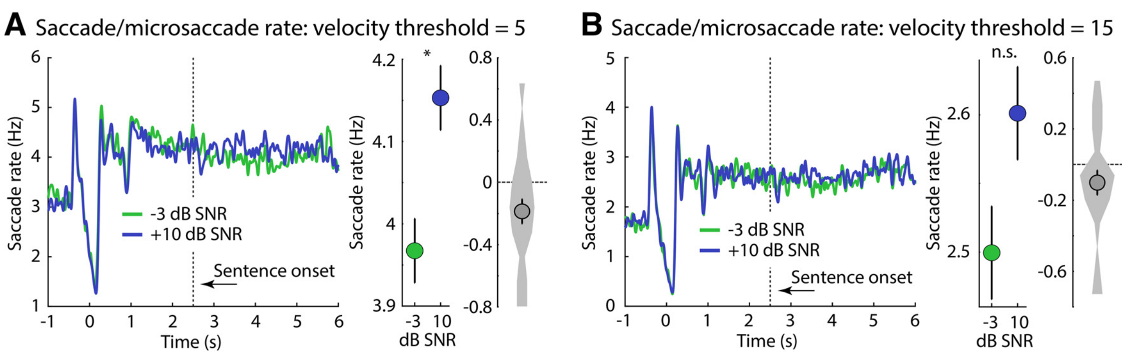

Figure 6. Results for saccade/microsaccade rate. ***A***, Left, Saccade/microsaccade rate time courses for velocity threshold of five times the median eye-movement velocity SD. The vertical dashed line marks the sentence onset. Middle, Mean saccade/microsaccade rate (across the 2.5–4.9 s time window; sentence onset at 2.5 s and average sentence offset at 4.9 s). Error bars indicate SEM (removal of between-participant variance; Masson and Loftus, 2003). Right, Difference in fixation duration between the two SNR levels (−3 dB SNR minus +10 dB SNR). The violin plot reflects the histogram of individual data points. The dot reflects the mean across participants. ***B***, Same as in ***A*** for saccade/microsaccade rate calculated using a velocity threshold of 15 times the median eye-movement velocity SD; $*p \leq 0.05$; n.s.

2021). The eye-movement metrics established in experiments 1 and 2 (i.e., fixation time and spatial gaze dispersion) do not require baseline normalization and may thus be uniquely sensitive to listening effort during continuous speech listening.

Participants

Twenty-three adults (median age, 25 years; age range, 18–34 years; 15 female, 6 male, 1 nonbinary; 1 person did not answer) participated in experiment 3. Data from one additional participant were recorded but excluded from analysis because >40% of their data were rejected during preprocessing. Twenty-one of the 23 participants were native English speakers, the other 2 participants were highly proficient non-native English speakers. The mean self-rated English skills of non-native English speakers on a 0 (poor) to 10 (very high) scale was 8.5.

Acoustic stimulation and procedure

Participants listened to two ~10 min stories from the storytelling podcast *The Moth* [https://themoth.org/; "Nacho Challenge" by Omar Qureshi (11 min), "Family Trees Can Be Dangerous" by Paul Nurse (10 min)]. *The Moth* consists of people telling stories about interesting life events, and the stories are highly enjoyable and absorbing (Herrmann and Johnsrude, 2020b; Irsik et al., 2022b). Each participant listened to one of the stories in the original temporal order (referred to as "intact") and the other story in scrambled order; that is, the order of phrases/sentences was shuffled (referred to as "scrambled"). A scrambled story was included as a control story that resembles short, disconnected sentences for which we expected story comprehension, listening motivation, and effort investment to be low. Scrambled stories were obtained using a custom MATLAB script as follows. Silence periods with a duration of at least 0.05 s were identified. The duration of speech snippets separated by the identified silence periods was calculated. Speech snippets with a duration of 1 s or longer were cut out at the center of the silence periods, and the order of snippets was shuffled. Story scrambling was calculated for each participant uniquely. Which of the two *The Moth* stories was intact versus scrambled was counterbalanced across participants. The order in which the intact versus the scrambled story was presented was also counterbalanced across participants.

Each story was masked by 12-talker background babble (Bilger, 1984). The SNR between the speech signal and the 12-talker babble changed every 28 s to one of five SNR levels (+16, +11, +6, +1, −4 dB SNR; Irsik et al.,2022a,b), corresponding to about a range of 95 to 50% of intelligible words (Irsik et al., 2022b). Each SNR was manipulated by adjusting the dB level of both the story and the masker. This ensured that the overall sound level remained constant across SNRs and throughout a story and that the overall level was similar for both stories. Each story started and ended with the +16-dB SNR level to enable participants to clearly hear the beginning and end of the story. Each SNR level was presented four times, except for the +16 dB SNR level, which was presented six times (four times plus beginning and end). The SNR transitioned smoothly from one level to the other over a duration of 1 s. The order of SNR levels was randomized such that a particular SNR could not be heard twice in succession, and that SNR would maximally change by two levels. For each participant, SNR levels were randomized uniquely, but the same randomization was used for the intact and the scrambled story.

The two stories were presented in two separate blocks, and participants took a break between blocks. Each story was presented continuously for the ~10 min duration without any silence periods or interruptions (SNR levels seamlessly transitioned). That is, in contrast to previous work and experiments 1 and 2 (Wendt et al., 2018; Winn et al., 2018; Zhao et al., 2019b; Fiedler et al., 2021), no silent or nonspeech baseline period nor unique disconnected trials were available in our naturalistic story-listening paradigm.

After a story ended, participants answered 10 comprehension questions about the story. For each comprehension question, participants were asked to select the correct answer of four multiple-choice options.

Visual stimulation

To facilitate eye movements, we adopted an incidental multiple-object movement display (Cavanagh and Alvarez, 2005; Alvarez and Franconeri, 2007; Scholl, 2009; Herrmann and Johnsrude, 2018a, b) that was concurrently presented with the spoken stories. To this end, 16 dots [dot diameter, 1.2 cm (0.9°)] were presented and moved on the screen. The presentation of dots was constrained to a display frame of 20.6 cm in width (15.6°) and 19.4 cm in height (14.7°) centered on the screen and highlighted for the participants by a gray frame on a gray background (see Fig. 7*B*). Dots never moved outside the display frame and never overlapped during movements; dots moved ~3.7 cm/s (2.8°/s). The locations of the 16 dots were set to new randomly selected locations every 3–5 s to facilitate eye movements and overcome the technical challenge of displaying continuous dot movements for the ~10 min story duration, which exceeded the working memory capacity of the laptop running the experimental procedures.

Critically, participants were instructed that their task was to comprehend the story so that they would be able to answer the comprehension questions following the story presentation. Hence, participants did not need to perform a task on the multiple-object movement display. Participants were instead instructed to look at the screen in whatever way they wanted (Johansson et al., 2006, 2011, 2012).

Analysis of behavioral data

Responses to comprehension questions were coded as correct or incorrect. A mean story-comprehension score was calculated as the proportion of correct responses across the 10 comprehension questions, separately for the intact and the scrambled stories. A dependent-samples *t* test was calculated to compare the proportion of correct responses between the intact and the scrambled story.

Analysis of eye-movement and pupil-area data

For each story, the ~10 min time courses of the pupil area and *x* and *y* coordinates were used for analysis (preprocessing was as described above in Materials and Methods). Specifically, fixation duration, spatial gaze dispersion, and pupil area were obtained for each time point over the duration of a story.

To obtain a fixation-duration time course of each story, the number of continuous presamples and postsamples for which the *x* and **y** eye coordinates remained in a given location was calculated separately for each time point. The gaze dispersion time course for each story was obtained by calculating gaze dispersion across a 1 s time window centered on each time point. For the pupil area time courses, the data following preprocessing were used (for details see above, Materials and Methods).

For the analysis of the SNR effects, fixation duration, gaze dispersion, and pupil area were averaged across time points, separately for each of the five SNR levels and the two story types (intact, scrambled). Analyses of how the different measures evolved over time following an SNR change could not be analyzed because too few repetitions per SNR were available (four for +11, +6, +1, −4 dB SNR, six for +16 dB SNR). Saccade/microsaccade rate was also not calculated for story-listening data because the method to estimate what counts as a saccade/microsaccade has been developed for trial-based and not continuous data (Engbert and Kliegl, 2003; Engbert, 2006; Widmann et al., 2014; Kadem et al., 2020).

For each participant and story type, a linear function was fit separately to fixation duration, gaze dispersion, and pupil area as a function of SNR levels. The resulting slope (linear coefficient) was tested against zero using a one-sample *t* test to test whether there was a significant relation between SNR levels and the eye metrics. A dependent-samples *t* test was calculated to compare the slopes between the intact and the scrambled story.

Results

The proportion of correctly answered comprehension questions was greater for the intact story compared with the scrambled story, as expected ($t_{(22)} = 8.43$, $p = 2.4 \cdot 10^{-8}$, $d = 1.758$; Fig. 7*C*).

For fixation duration, the slope of a linear function fit relating SNR to fixation duration showed that fixation duration decreased with increasing SNR for the intact story ($t_{(22)} = -2.660$, $p = 0.014$, $d = 0.555$) but not for the scrambled story ($t_{(22)} = 0.440$, $p = 0.664$, $d = 0.092$; Fig. 8*A*; difference between slopes, $t_{(22)} = 2.043$, $p = 0.053$, $d = 0.426$). In fact, the mean fixation duration (across SNRs) for the scrambled story did not differ from the fixation duration for the most favorable SNR of the intact story (+16 dB; $t_{(22)} = 0.685$, $p = 0.500$, $d = 0.143$) but was smaller than the fixation duration for the most unfavorable SNR (−4 dB; $t_{(22)} = 2.672$, $p = 0.014$, $d = 0.557$). In other words, fixation duration during listening to the scrambled story mirrored the fixation duration when speech comprehension was easy during listening to the intact story, consistent with hypothesis that individuals may not invest cognitively as much while listening to a scrambled story.

Results for spatial gaze dispersion mirrored those for fixation duration. Gaze dispersion increased with increasing SNR for the intact story ($t_{(22)} = 4.250$, $p = 3.3 \cdot 10^{-4}$, $d = 0.886$) but not for

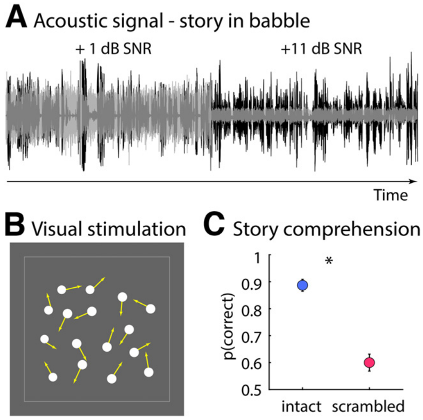

Figure 7. Stimulation and behavioral data for experiment 3. ***A***, Sample snippet of the engaging, spoken story masked by background babble. The SNR between the speech signal and the background babble changed every 28 s to one of five SNR levels (−4, +1, +6, +11, +16 dB SNR). ***B***, Moving-dot display presented while participants listened to the spoken story. Sixteen dots smoothly moved on the screen. No task was required using the moving-dot display. ***C***, Proportion of correctly answered story comprehension questions for the intact and scrambled story (the temporal order of phrases/sentences were shuffled within the scrambled story); $^{*}p < 0.05$.

the scrambled story ($t_{(22)} = 1.609$, $p = 0.122$, $d = 0.336$; difference between slopes, $t_{(22)} = 2.147$, $p = 0.043$, $d = 0.448$). Moreover, mean gaze dispersion for the scrambled story (across SNRs) was not different from gaze dispersion at the most favorable SNR of the intact story (+16 dB; $t_{(22)} = 0.257$, $p = 0.800$, $d = 0.054$) but was significantly greater compared with the least favorable SNR of the intact story (−4 dB; $t_{(22)} = 3.102$, $p = 0.005$, $d = 0.647$). These data thus suggest again that individuals' eyes moved as much during the scrambled story as they did during the easiest SNR while listening to the intact story, indicating that listeners did not invest cognitively while listening to the scrambled story.

In contrast to previous and current pupil data recorded during sentence listening (experiments 1 and 2; Zekveld et al., 2010; Wendt et al., 2016; Kadem et al., 2020), we found no linear relation between SNR and pupil area for the intact story ($t_{(22)} = 0.014$, $p = 0.989$, $d = 0.003$; Fig. 9). Some previous works have shown a nonlinear relationship of the pupil response as a function of SNR, such that the pupil area is largest for moderate SNRs and smaller for unfavorable and favorable SNRs (Ohlenforst et al., 2017; Wendt et al., 2018). However, the pupil area in experiment 3 showed the opposite pattern. There was also an unexpected linear increase in pupil area with increasing SNR for the scrambled story ($t_{(22)} = 2.121$, $p = 0.045$, $d = 0.442$; Fig. 9; no difference between the intact and the scrambled story, $t_{(22)} = 1.525$, $p = 0.142$, $d = 0.318$). The absence of a linear SNR effect on pupil area during engaging story listening may be related to the incidental moving-dot display, reducing sensitivity, and/or because of the omission of pupil baseline normalization, which is not feasible for continuous speech listening over several minutes. The increase in pupil area with increasing SNR for the scrambled story perhaps reflects arousal (Bradshaw, 1967; Bradley et al., 2008; Mathôt, 2018; Ayasse and Wingfield, 2020) when speech becomes unmasked for low-engaging materials.

Summary

Experiment 3 shows that individuals' eye movements decrease when speech is masked by background noise during listening to an engaging, continuous story. The sensitivity of eye movements to listening challenges induced by background masking was specific to the interesting, engaging story and was absent for a less engaging, scrambled story. Results of experiment 3 further led to unexpected pupillometry results during continuous story listening, for example, a larger pupil size for more favorable SNRs. The current data may indicate challenges with pupillometric measures in more naturalistic speech-listening paradigms and suggest that other eye-based metrics such as fixation duration and gaze dispersion could provide an alternative.

Data availability

Data are available at https://osf.io/pb4dr/.

Discussion

Pupillometry is the most used objective tool to assess listening effort but has multiple disadvantages. The current study explored a new way to assess listening effort based on eye movements. In three experiments, we show that listeners' eye movements decrease when listening to masked speech is effortful, as indicated by increased fixation duration and decreased spatial gaze dispersion. We demonstrate this effort-related reduction in eye movements during free viewing and incidental object tracking, as well as for simple sentences and naturalistic, continuous stories. Pupillometry was not sensitive to listening effort during story listening (only during isolated sentence listening), highlighting the challenges with pupillometric measures for the assessments of listening effort in naturalistic speech-listening paradigms. Our results reveal a critical link between eye movements and cognitive load that can be leveraged to measure listening effort.

Reduced eye movements as a measure of cognitive load

The current work demonstrates that eye movements reduce during difficult listening. Consistently, a few studies have shown that individuals avert gaze (Glenberg et al., 1998), reduce object-tracking eye movements (Lipton et al., 1980; Hutton and Tegally, 2005; Kosch et al., 2018), and decrease saccades (Walter and Bex, 2021) when memory load is high compared with low. It thus appears that different cognitively challenging tasks affect oculomotor function such that eye movements decrease during periods of high cognitive load.

We also observed that saccades/microsaccades are sensitive to speech masking, but this appeared to depend on the threshold that defined what counts as a saccade/microsaccade in inconsistent ways across experiment 1 and 2 (Figs. 3, 6). The saccade/microsaccade rate was also not sensitive to the different viewing conditions (free vs fixation) in experiment 1, whereas fixation duration and gaze dispersion were. Some previous research did not find that listening effort affects saccades/microsaccades, but different thresholds were not tested (Kadem et al., 2020). Other very recent work indicates microsaccades can be modulated by listening challenges (Contadini-Wright et al., 2023). Saccades/microsaccades occur relatively infrequently ∼1–3 times per second (Martinez-Conde et al., 2009, 2013; Pierce et al., 2019), which may make saccades/microsaccades less sensitive to listening effort. The eye-movement measures proposed in the current study—fixation duration and gaze dispersion—do not focus on a specific type of eye movement but instead leverage all x and y data within a time window as a basis for calculations, possibly leading to increased sensitivity. Nevertheless, given the inconsistent effect of speech masking on saccades/microsaccades, the current data may suggest that nonsaccadic eye movements,

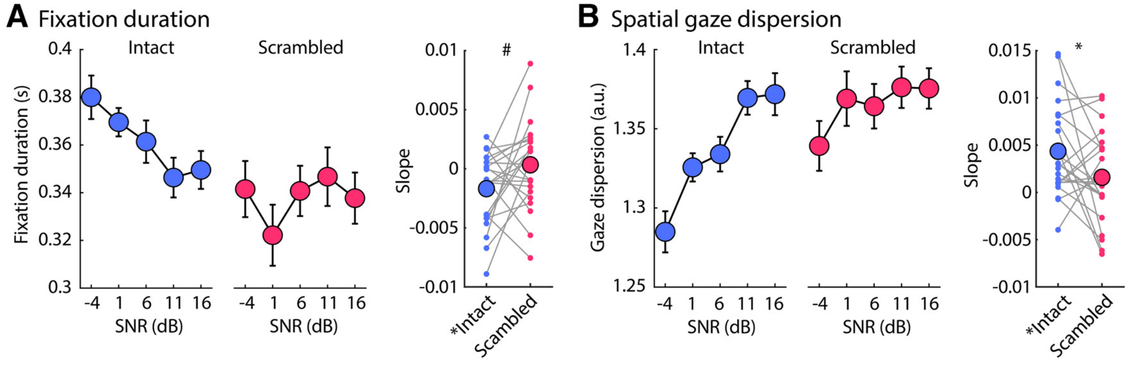

Figure 8. Eye movements decrease when story listening is challenging. ***A***, Left, Fixation duration for the intact and the scrambled story for each SNR level. The error bars indicate SEM (removal of between-participants variance; Masson and Loftus, 2003). Right, Plot shows the slope from a linear function fit, reflecting the linear relation between SNR levels and fixation duration. Small dots reflect data points from individual participants. The asterisk in front of the condition label indicates a significant difference from zero ($p \leq 0.05$). ***B***, Same as in A for spatial gaze dispersion; $*p \leq 0.05$, $\#p \leq 0.1$.

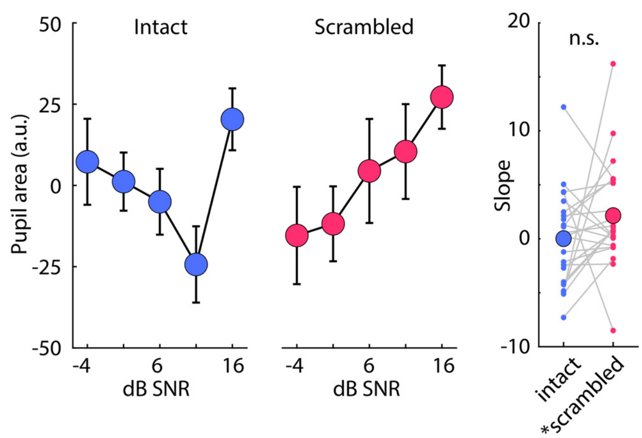

Figure 9. Pupil area during story listening. Left, Pupil area for the intact and the scrambled story for each SNR level. The error bars indicate SEM (removal of between-participants variance; Masson and Loftus, 2003). Right, Plot shows the slope from a linear function fit, reflecting the linear relation between SNR levels and pupil area. The asterisk in front of the condition label indicates a significant difference from zero ($p \leq 0.05$); n.s.

possibly smooth eye movements, contribute to the reduction in eye movements during effortful speech listening.

Different sensitivity of pupil area versus eye-movement metrics to listening effort

We show that pupil area increases with increasing effort induced by masked speech during sentence listening (Figs. 2*C*, 5*D*), in line with a large body of previous work (Zekveld et al., 2010; Kuchinsky et al., 2013; Zekveld et al., 2014; Zekveld and Kramer, 2014; Winn et al., 2015; Wendt et al., 2016; Winn, 2016; Wendt et al., 2017; Zekveld et al., 2019; Kadem et al., 2020; Seifi Ala et al., 2020; Winn and Teece, 2021; Zhang et al., 2022; Neagu et al., 2023). The current results further show that the masking-related increase in pupil area is also present when fixation on a stationary point is not required (Kraus et al., 2022). Fixation to a stationary point reduces the impact of SNR on speech comprehension (Fig. 1) and may impair memory and mental imagery for spoken speech (Johansson et al., 2012). The observation that pupil area is sensitive to listening effort during free viewing (Fig. 2*C*) and incidental object tracking (Fig. 5*D*) perhaps suggests that strict fixation to assess listening effort with pupillometry is not required.

Pupil area was not sensitive to effort-related speech masking when individuals listened to continuous, engaging stories (Fig. 9). In contrast, fixation duration increased, and spatial gaze dispersion decreased with increased speech masking during both sentence and story listening (Figs. 2, 5, 8), emphasizing the potential of using eye movements to assess listening effort for naturalistic speech listening. Previous work using continuous speech materials that consisted of 30 s passages, rather than the ~2–3 s sentences typically used, observed a larger pupil size for less compared with more favorable speech-masking levels (Seifi Ala et al., 2020; Fiedler et al., 2021; for a study using non-speech stimuli, Zhao et al., 2019b). Pupil size in these studies was normalized to a prespeech baseline, which is also commonly done in sentence-listening paradigms (Winn et al., 2018; Zekveld et al., 2018). Baseline normalization is not attainable for ~10 min continuous stories because no neutral, speech-devoid time period is available. The absence of baseline normalization in our story-listening experiment may have contributed to the absence of an SNR effect on pupil area for the engaging story. The eye-movement metrics proposed here—fixation duration and spatial gaze dispersion—do not require baseline normalization and may thus be uniquely sensitive to speech masking during story listening.

Our analyses show that changes in pupil area and changes in eye-movement metrics do not significantly correlate. Whereas the current study was not designed to specifically examine inter-individual variations—that is, correlation analyses may require a higher number of participants (Bossier et al., 2020; Grady et al., 2021)—the absence of a correlation between pupil-area and eye-movement changes may imply distinct processes or mechanisms. (Although, note that the direction of the relation was as expected.) Previous works have also failed to find a significant relationship between indices of listening effort, for example, among pupil area, neural oscillation, heart rate, and skin conductance (Miles et al., 2017; Strand et al., 2018; Alhanbali et al., 2019; Seifi Ala et al., 2020; Kraus et al., 2022). The absence of a relationship between listening-effort measures might be because of different neural mechanisms that tap into different aspects of effort (Strand et al., 2018; Herrmann and Johnsrude, 2020a; Strand et al., 2020). Given that the number of participants in the current study may be on the lower end to make concluding inferences about the correlation between measures, we recommend further examination in future research.

Neural mechanisms of reduced eye movements during challenging listening

That eye movements reduce with increasing listening effort aligns with work in nonhuman mammals, showing increased activity in auditory cortex during periods of reduced movements (Schneider et al., 2014; McGinley et al., 2015; Schneider and Mooney, 2015, 2018; O'Connell et al., 2020), and heightened neuronal excitability in auditory cortex in the absence of eye movements (O'Connell et al., 2020). Reductions in any (eye) movements may thus support auditory processing, including speech perception, by enhancing auditory-system sensitivity. The decrease in eye movements may further reduce visual, proprioceptive, and other inputs that may distract cognitively from listening.

Generation and regulation of eye movements relies on a network of neural circuits that involves cortical and subcortical brain structures, such as the visual cortex, prefrontal cortex, posterior parietal cortex, frontal and supplementary eye fields, anterior cingulate cortex, cerebellum, thalamus, basal ganglia, and superior colliculus (Sparks, 2002; Pierrot-Deseilligny et al., 2004; Pierce et al., 2019). Some of these regions overlap with the network that regulates the pupil size (Wang et al., 2012; Joshi and Gold, 2020; Wang and Munoz, 2021; Burlingham et al., 2022), highlighting the potential for shared mechanisms of effort-related changes in pupil area and eye movements.

Which regions drive the reduction in eye movements during challenging listening is unclear. We speculate that cognitive control regions, such as the prefrontal, cingulate, and parietal cortices (Cole and Schneider, 2007; Braver, 2012; Niendam et al., 2012), may influence structures that initiate eye movements, such as the frontal and supplementary eye fields and the superior colliculus (Pierce et al., 2019). However, the degree of network activation has been shown to depend on whether eye movements are automatically triggered or volitionally initiated (Pierce et al., 2019). As such, the specific involvement of brain regions may depend on whether listeners consciously choose to reduce eye movements when listening becomes effortful or whether eye movements decrease automatically. Although eye movements are under cognitive control, the observation of increased fixation duration with decreasing SNR while listeners fixate on a stationary point (Fig. 2) perhaps suggests involuntary contributions. The question of whether the reduction of eye movements is deliberate versus involuntary also has implications for the type of effort the reduced eye movements index. A deliberate reduction implies that the listener experiences listening effort first and subsequently reduces eye movements, whereas an involuntary process may index cognitive resource recruitment rather than the experience of effort (Lemke and Besser, 2016; Herrmann and Johnsrude, 2020a; Strand et al., 2020, for important discussions on experienced vs exerted effort).

Conclusions

Pupillometry is the most used objective tool to assess listening effort but has several disadvantages. Building on cognitive and neurophysiological work, the current study explores a new way to assess listening effort through eye movements. Here, we examine the hypothesis that eye movements decrease when speech listening becomes effortful. Consistent with this hypothesis, we demonstrate that fixation duration increases, and spatial gaze dispersion decreases with increasing speech masking. Eye movements decrease when speech comprehension is effortful during free viewing and object tracking, as well as for sentence and story listening. Pupillometry was sensitive to speech masking only during sentence listening, but not during story listening, highlighting the challenges with pupillometric measures for the assessments of listening effort during naturalistic speech listening. Our results reveal a critical link between eye movements and cognitive load during speech comprehension and provide the foundation for a novel measure of listening effort that has the potential to be applicable in a wide range of contexts.

References

Alhanbali S, Dawes P, Millman RE, Munro KJ (2019) Measures of listening effort are multidimensional. Ear Hear 40:1084–1097.

Alvarez GA, Franconeri SL (2007) How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. J Vis 7 (13):14, 1– 10.

Ayasse ND, Wingfield A (2018) A tipping point in listening effort: effects of linguistic complexity and age-related hearing loss on sentence comprehension. Trends Hear 22:2331216518790907.

Ayasse ND, Wingfield A (2020) Anticipatory baseline pupil diameter is sensitive to differences in hearing thresholds. Front Psychol 10:2947.

Bilger RC (1984) Manual for the clinical use of the revised SPIN test. Champaign, IL: University of Illinois.

Bohanek JG, Fivush R, Zaman W, Lepore CE, Merchant S, Duke MP (2009) Narrative interaction in family dinnertime conversations. Merrill Palmer Q (Wayne State Univ Press) 55:488–515.

Bossier H, et al. (2020) The empirical replicability of task-based fMRI as a function of sample size. Neuroimage 212:116601.

Bradley MM, Miccoli L, Escrig MA, Lang PJ (2008) The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology 45:602–607.

Bradshaw J (1967) Pupil size as a measure of arousal during information processing. Nature 216:515–516.

Braver TS (2012) The variable nature of cognitive control: a dual mechanisms framework. Trends Cogn Sci 16:106–113.

Brisson J, Mainville M, Mailloux D, Beaulieu C, Serres J, Sirois S (2013) Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. Behav Res Methods 45:1322–1331.

Burlingham CS, Mirbagheri S, Heeger DJ (2022) A unified model of the task-evoked pupil response. Sci Adv 8:eabi9979.

Carolan PJ, Heinrich A, Munro KJ, Millman RE (2022) Quantifying the effects of motivation on listening effort: a systematic review and meta-analysis. Trends Hear 26:23312165211059982.

Cavanagh P, Alvarez GA (2005) Tracking multiple targets with multifocal attention. Trends Cogn Sci 9:349–354.

Cole MW, Schneider W (2007) The cognitive control network: integrated cortical regions with dissociable functions. Neuroimage 37:343–360.

Contadini-Wright C, Magami K, Mehta N, Chait M (2023) Pupil dilation and microsaccades provide complementary insights into the dynamics of arousal and instantaneous attention during effortful listening. J Neurosci 43:4856–4866.

Dimitrijevic A, Smith ML, Kadis DS, Moore DR (2017) Cortical alpha oscillations predict speech intelligibility. Front Hum Neurosci 11:88.

Engbert R (2006) 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.

Farahani M, Parsa V, Herrmann B, Kadem M, Johnsrude IS, Doyle PC (2020) An auditory-perceptual and pupillometric study of vocal strain and listening in adductor spasmodic dysphonia. Applied Sciences 10:5907.

Feder K, Michaud D, Ramage-Morin P, McNamee J, Beauregard Y (2015) Prevalence of hearing loss among Canadians aged 20 to 79: audiometric results from the 2012/2013 Canadian Health Measures Survey. Health Rep 26:18–25.

Fiedler L, Seifi Ala T, Graversen C, Alickovic E, Lunner T, Wendt D (2021) Hearing aid noise reduction lowers the sustained listening effort during continuous speech in noise—a combined pupillometry and EEG study. Ear Hear 42:1590–1601.

Fink L, Simola J, Tavano A, Lange E, Wallot S, Laeng B (2023) From pre-processing to advanced dynamic modeling of pupil data. Behav Res Methods. Advance online publication. Retrieved July 24, 2023. https://doi.org/10.3758/s13428-023-02098-1.

Fivush R, Bohanek JG, Zaman W (2011) Personal and intergenerational narratives in relation to adolescents' well-being. New Dir Child Adolesc Dev 2011:45–57.

Fukushima K (2003) Frontal cortical control of smooth-pursuit. Curr Opin Neurobiol 13:647–654.

Gatehouse S, Noble W (2004) The speech, spatial and qualities of hearing scale (SSQ). Int J Audiol 43:85–99.

Glenberg AM, Schroeder JL, Robertson DA (1998) Averting the gaze disengages the environment and facilitates remembering. Mem Cognit 26:651–658.

Goman AM, Lin FR (2016) Prevalence of hearing loss by severity in the United States. Am J Public Health 106:1820–1822.

Grady CL, Rieck JR, Nichol D, Rodrigue KM, Kennedy KM (2021) Influence of sample size and analytic approach on stability and interpretation of brain-behavior correlations in task-related fMRI data. Hum Brain Mapp 42:204–219.

Hayes TR, Petrov AA (2016) Mapping and correcting the influence of gaze position on pupil size measurements. Behav Res Methods 48:510–527.

Herrmann B, Johnsrude IS (2018a) Attentional state modulates the effect of an irrelevant stimulus dimension on perception. J Exp Psychol Hum Percept Perform 44:89–105.

Herrmann B, Johnsrude IS (2018b) Neural signatures of the processing of temporal patterns in sound. J Neurosci 38:5466–5477.

Herrmann B, Johnsrude IS (2020a) A model of listening engagement (MoLE). Hear Res 397:108016.

Herrmann B, Johnsrude IS (2020b) Absorption and enjoyment during listening to acoustically masked stories. Trends Hear 24: 23312165211031751.

Hutton SB, Tegally D (2005) The effects of dividing attention on smooth pursuit eye tracking. Exp Brain Res 163:306–313.

Institute of Electrical and Electronics Engineers (1969) IEEE recommended practice for speech quality measurements. IEEE Trans Audio Electroacoust 17:225–246.

Irsik VC, Johnsrude IS, Herrmann B (2022a) Age-related deficits in dip-listening evident for isolated sentences but not for spoken stories. Sci Rep 12:5898.

Irsik VC, Johnsrude IS, Herrmann B (2022b) Neural activity during story listening is synchronized across individuals despite acoustic masking. J Cogn Neurosci 34:933–950.

Jefferson G (1978) Sequential aspects of storytelling in conversation. In: Studies in the organization of conversational interaction (Schenkein JIM, ed), pp 219–248. New York: Academic.

Johansson R, Holsanova J, Holmqvist K (2006) Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. Cogn Sci 30:1053–1079.

Johansson R, Holsanova J, Homqvist K (2011) The dispersion of eye movements during visual imagery is related to individual differences in spatial imagery ability. Cogn Sci 33:1200–1205.

Johansson R, Holsanova J, Dewhurst R, Holmqvist K (2012) Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. J Exp Psychol Hum Percept Perform 38:1289–1314.

Joshi S, Gold JI (2020) Pupil size as a window on neural substrates of cognition. Trends Cogn Sci 24:466–480.

Kadem M, Herrmann B, Rodd JM, Johnsrude IS (2020) Pupil dilation is sensitive to semantic ambiguity and noise. Trends Hear 24:2331216520964068.

Kinley I, Levy Y (2022) PuPl: an open-source tool for processing pupillometry data. Behav Res 54:2046–2069.

Knapen T, de Gee JW, Brascamp J, Nuiten S, Hoppenbrouwers S, Theeuwes J (2016) Cognitive and ocular factors jointly determine pupil responses under equiluminance. PLoS One 11:e0155574.

Koelewijn T, Zekveld AA, Lunner T, Kramer SE (2018) The effect of reward on listening effort as reflected by the pupil dilation response. Hear Res 367:106–112.

Kosch T, Hassib M, Woźniak PW, Buschek D, Alt F (2018) Your eyes tell: leveraging smooth pursuit for assessing cognitive workload. Paper presented at the CHI Conference on Human Factors in Computing Systems, Montreal, April.

Kraus F, Tune S, Obleser J, Herrmann B (2022) Neural α oscillations and pupil size differentially index cognitive demand under competing audiovisual task conditions. J Neurosci 43:4352–4364.

Krueger M, Schulte M, Brand T, Holube I (2017) Development of an adaptive scaling method for subjective listening effort. J Acoust Soc Am 141:4680–4693.

Kuchinsky SE, Ahlstrom JB, Vaden KI Jr, Cute SL, Humes LE, Dubno JR, Eckert MA (2013) Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. Psychophysiology 50:23–34.

Larsby B, Hällgren M, Lyxell B, Arlinger S (2005) Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. Int J Audiol 44:131–143.

Lemke U, Besser J (2016) Cognitive load and listening effort: concepts and age-related considerations. Ear Hear 37 Suppl 1:77S–84S.

Lipton RB, Frost LA, Holzman PS (1980) Smooth pursuit eye movements, schizophrenia, and distraction. Percept Mot Skills 50:159–167.

Mackersie CL, MacPhee IX, Heldt EW (2015) Effects of hearing loss on heart rate variability and skin conductance measured during sentence recognition in noise. Ear Hear 36:145–154.

Martinez-Conde S, Macknik SL, Troncoso XG, Hubel DH (2009) Microsaccades: a neurophysiological analysis. Trends Neurosci 32:463–475.

Martinez-Conde S, Otero-Millan J, Macknik SL (2013) The impact of microsaccades on vision: towards a unified theory of saccadic function. Nat Rev Neurosci 14:83–96.

Masson MEJ, Loftus GR (2003) Using confidence intervals for graphically based data interpretation. Can J Exp Psychol 57:203–220.

Mathôt S (2018) Pupillometry: psychology, physiology, and function. J Cogn 1:16.

McGinley MJ, David SV, McCormick DA (2015) Cortical membrane potential signature of optimal states for sensory signal detection. Neuron 87:179–192.

McIntire LK, McIntire JP, McKinley RA, Goodyear C (2014) Detection of vigilance performance with pupillometry. Paper presented at the Symposium on Eye Tracking Research and Applications, Safety Harbor, Florida, March.

Miles K, McMahon C, Boisvert I, Ibrahim R, de Lissa P, Graham P, Lyxell B (2017) Objective assessment of listening effort: coregistration of pupillometry and EEG. Trends Hear 21:33121651770639.

Missal M, Heinen SJ (2017) Stopping smooth pursuit. Phil Trans R Soc B 372:20160200.

Moore TM, Picou EM (2018) A potential bias in subjective ratings of mental effort. J Speech Lang Hear Res 61:2405–2421.

Mullen MK, Yi S (1995) The cultural context of talk about the past: implications for the development of autobiographical memory. Cogn Dev10:407–419.

Neagu M-B, Kressner AA, Relaño-Iborra H, Bækgaard P, Dau T, Wendt D (2023) Investigating the reliability of pupillometry as a measure of individualized listening effort. Trends Hear 27:23312165231153288.

Niendam TA, Laird AR, Ray KL, Dean YM, Glahn DC, Carter CS (2012) Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. Cogn Affect Behav Neurosci 12:241–268.

O'Connell MN, Barczak A, McGinnis T, Mackin K, Mowery T, Schroeder CE, Lakatos P (2020) The role of motor and environmental visual rhythms in structuring auditory cortical excitability. iScience 23:101374.

Ohlenforst B, Zekveld AA, Lunner T, Wendt D, Naylor G, Wang Y, Versfeld NJ, Kramer SE (2017) Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. Hear Res 351:68–79.

Ono S (2015) The neuronal basis of on-line visual control in smooth pursuit eye movements. Vision Res 110:257–264.

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, Levitt H (2012) Speech comprehension training and auditory and cognitive processing in older adults. Am J Audiol 21:351–357.

Pichora-Fuller MK, Schneider BA, Daneman M (1995) How young and old adults listen to and remember speech in noise. J Acoust Soc Am 97:593–608.

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 Suppl 1:5S–27S.

Pierce JE, Clementz BA, McDowell JE (2019) Saccades: fundamentals and neural mechanisms. In: Eye movement research: an introduction to its scientific foundations and applications. (Klein C, Ettinger U, eds), pp 11–71. Cham, Switzerland: Springer.

Pierrot-Deseilligny C, Milea D, Müri RM (2004) Eye movement control by the cerebral cortex. Curr Opin Neurol 17:17–25.

Poletti M, Rucci M (2016) A compact field guide to the study of microsaccades: challenges and functions. Vision Res 118:83–97.

Rodd JM, Longe OA, Randall B, Tyler LK (2010a) The functional organisation of the fronto-temporal language system: evidence from syntactic and semantic ambiguity. Neuropsychologia 48:1324–1335.

Rodd JM, Johnsrude IS, Davis MH (2010b) The role of domain-general frontal systems in language comprehension: evidence from dual-task interference and semantic ambiguity. Brain Lang 115:182–188.

Schneider DM, Mooney R (2015) Motor-related signals in the auditory system for listening and learning. Curr Opin Neurobiol 33:78–84.

Schneider DM, Mooney R (2018) How movement modulates hearing. Annu Rev Neurosci 41:553–572.

Schneider DM, Nelson A, Mooney R (2014) A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature 513:189–194.

Scholl BJ (2009) What have we learned about attention from multiple object tracking (and vice versa)? In: Computation, cognition, and Pylyshyn (Dedrick D, Trick L, eds), pp 49–78. Cambridge, MA: MIT.

Seifi Ala T, Graversen C, Wendt D, Alickovic E, Whitmer WM, Lunner T (2020) An exploratory study of EEG alpha oscillation and pupil dilation in hearing-aid users during effortful listening to continuous speech. PLoS One 15:e0235782.

Sparks DL (2002) The brainstem control of saccadic eye movements. Nat Rev Neurosci 3:952–964.

Strand JF, Brown VA, Merchant MB, Brown HE, Smith J (2018) Measuring listening effort: convergent validity, sensitivity, and links with cognitive and personality measures. J Speech Lang Hear Res 61:1463–1486.

Strand JF, Ray L, Dillman-Hasso NH, Villanueva J, Brown VA (2020) Understanding speech amid the jingle and jangle: recommendations for improving measurement practices in listening effort research. Audit Percept Cogn 3:169–188.

Suzuki Y, Minami T, Laeng B, Nakauchi S (2019) Colorful glares: effects of colors on brightness illusions measured with pupillometry. Acta Psychol (Amst) 198:102882.

Thurman SM, Cohen Hoffing RA, Madison A, Ries AJ, Gordon SM, Touryan J (2021) “Blue sky effect”: contextual influences on pupil size during naturalistic visual search. Front Psychol 12:748539.

Walter K, Bex P (2021) Cognitive load influences oculomotor behavior in natural scenes. Sci Rep 11:12405.

Wang C-A, Munoz DP (2021) Coordination of pupil and saccade responses by the superior colliculus. J Cogn Neurosci 33:919–932.

Wang C-A, Boehnke SE, White BJ, Munoz DP (2012) Microstimulation of the monkey superior colliculus induces pupil dilation without evoking saccades. J Neurosci 32:3629–3636.

Wendt D, Dau T, Hjortkjær J (2016) Impact of background noise and sentence complexity on processing demands during sentence comprehension. Front Psychol 7:345.

Wendt D, Hietkamp RK, Lunner T (2017) impact of noise and noise reduction on processing effort: a pupillometry study. Ear Hear 38:690–700.

Wendt D, Koelewijn T, Książek P, Kramer SE, Lunner T (2018) Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. Hear Res 369:67–78.

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 20:2331216516669723.

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 KH (2021) Listening effort is not the same as speech intelligibility score. Trends Hear 25:23312165211027688.

Winn MB, Edwards JR, Litovsky RY (2015) The impact of auditory spectral resolution on listening effort revealed by pupil dilation. Ear Hear 36: e153–e165.

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.

Wöstmann M, Herrmann B, Wilsch A, Obleser J (2015) Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. J Neurosci 35:1458–1467.

Zekveld AA, Kramer SE (2014) Cognitive processing load across a wide range of listening conditions: insights from pupillometry. Psychophysiology 51:277–284.

Zekveld AA, Kramer SE, Festen JM (2010) Pupil response as an indication of effortful listening: the influence of sentence intelligibility. Ear Hear 31:480–490.

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.

Zekveld AA, van Scheepen JAM, Versfeld NJ, Veerman ECI, Kramer SE (2019) Please try harder! The influence of hearing status and evaluative feedback during listening on the pupil dilation response, saliva-cortisol and saliva alpha-amylase levels. Hear Res 381:107768.

Zhang Y, Malaval F, Lehmann A, Deroche MLD (2022) Luminance effects on pupil dilation in speech-in-noise recognition. PLoS One 17:e0278506.

Zhao S, Wai Yum N, Benjamin L, Benhamou E, Yoneya M, Furukawa S, Dick F, Slaney M, Chait M (2019a) Rapid ocular responses are modulated by bottom-up-driven auditory salience. J Neurosci 39:7703–7714.

Zhao S, Bury G, Milne A, Chait M (2019b) Pupillometry as an objective measure of sustained attention in young and older listeners. Trends Hear 23: 2331216519887815.