Mobile Eye-Tracking Glasses Capture Ocular and Head Markers of Listening Effort

M. Eric Cui[1,2,*], Emilie Verno-Lavigne[1,2,*], Shreshth Saxena[3]

Lauren K. Fink[3], Björn Herrmann[1,2]

[1] *Rotman Research Institute, Baycrest Academy for Research and Education, North York, Canada, M6A 2E1*

[2] *Department of Psychology, University of Toronto, Toronto, Canada, M5S 1A1*

[3] *Department of Psychology, Neuroscience, & Behaviour, McMaster University, Hamilton, Canada L8S 4L8*

[*] *Co-first authors with equal contribution*

Correspondence: M. Eric Cui, mcui@research.baycrest.org, 3560 Kimel Family Building, Rotman Research Institute, Baycrest Academy for Research and Education, North York, Canada, M6A 2E1

Abstract: To extend the assessment of listening effort beyond a sound booth, we validated mobile eye-tracking glasses (Pupil Labs Neon) by comparing them to a stationary system (Eyelink DUO) in a controlled environment. We recorded eye movements, pupil size, and head movements from 26 young adults during a speech-in-noise task. When listening conditions became challenging, we observed reduced gaze dispersion and increased pupil sizes of similar magnitude from both devices, in addition to reduced head movements recorded solely by the mobile device. These findings suggest that mobile eye-trackers reliably capture listening effort, paving the path towards assessments in daily settings.

Keywords: listening effort, mobile eye-tracking glasses, pupillometry, eye movements, head movements

1. Introduction

Comprehending speech in noisy environments, such as a crowded restaurant, places high cognitive demand on attentional and memory resources, requiring listening effort (Herrmann & Johnsrude, 2020; Peelle, 2018; Pichora-Fuller et al., 2016a). Such effort is often considered an early indicator of hearing loss, making its objective assessment critical for clinical and research purposes (Helfer & Jesse, 2021; Herrmann & Johnsrude, 2020; Pichora-Fuller et al., 2016). To date, pupillometry is one of the most commonly used objective markers to listening effort, with numerous studies demonstrating that pupil size increases as listening becomes more demanding (Almaraz et al., 2018; Piquado et al., 2010; Van Der Wel & Van Steenbergen, 2018; Winn et al., 2018a; Zekveld et al., 2018). More recently, a separate line of work indicates that a reduction in eye movements may also index listening effort, potentially through pupil-independent mechanisms (Cui & Herrmann, 2023; Herrmann & Ryan, 2024). Assessing listening effort using eye movements offers a distinct advantage compared to pupillometry: eye-based metrics are less sensitive to environmental factors like ambient lighting, stimulus luminance, and baseline physiological fluctuations (Fink et al., 2024; Thurman et al., 2021; Winn et al., 2018a; Zekveld et al., 2018). This makes eye movements a promising candidate for assessing listening effort in more natural, real-world situations. However, both pupillometry and eye-movement research have mainly relied on stationary eye-trackers that require restricted head movements, limiting the ecological validity of these assessments.

Modern mobile eye-tracking glasses present an opportunity to overcome these limitations and move assessments towards real-world, out-of-lab contexts. These portable devices have been successfully used to track eye movements and pupil size across different age groups and hearing abilities (Chavant & Kapoula, 2022; Hidalgo et al., 2023; Perea Perez, 2023) for various auditory tasks, such as music listening and performing (Saxena et al., 2025; Vandemoortele et al., 2018), sound localization (Alemu et al., 2024), and speech-in-noise perception (Perea Perez, 2023). This evidence suggests that mobile eye-trackers could effectively measure listening effort. For example, promising correspondence between mobile eye-tracking glasses and a stationary EyeLink eye-tracker (the most commonly used system to achieve high temporal

and spatial precision) has been found for simple oculomotor tasks (Ehinger et al., 2019). However, it remains unclear whether relative measures, such as gaze dispersion (Cui & Herrmann, 2023; Herrmann et al., 2025; Herrmann & Ryan, 2024), a sensitive metric capturing changes in eye movements is robust enough to concurrent reduction in both spatial and temporal resolution.

Beyond expanding previously established listening effort measures from stationary to mobile devices, mobile eye-trackers provide additional unique metrics that may be relevant to listening effort. For instance, head movements (e.g., measured via inertial motion units embedded in head-mounted eye-trackers) may index cognitive load in a manner like eye movements, however, this has not been examined. Listeners reduce eye movements during difficult listening, possibly to limit distracting visual input (Cui & Herrmann, 2023). A similar pattern might occur for head movements, where a reduction in head movements might reduce distracting proprioceptive inputs, in turn freeing up cognitive resources for a listening task (Higgins et al., 2023). This would also be in line with research showing coupling between sensorimotor and cognitive systems (Dault et al., 2001; Craighero 2022). Integrating these markers could provide a comprehensive and ecologically valid approach to monitoring the dynamics of listening effort through mobile glasses.

The present study examines whether modern mobile eye-tracking glasses (Pupil Labs' Neon) are similarly sensitive to pupil-size and eye-movement markers of listening effort as a stationary eye-tracking system (EyeLink Duo). We also investigate whether head movements index listening effort. To this end, we concurrently measured pupil size, gaze dispersion, and head-movement variability with synchronized stationary and mobile devices concurrently while younger, normal-hearing adults performed a speech-comprehension task under easy and difficult speech-masking conditions.

2. Methods

2.1 Participants

Twenty-six younger adults (M = 25.75 years, SD = 5.25, range = 18-35 years; 17 females, 6 males, 1 agender, 2 non-binary) participated in the study. All self-reported being native English speakers, having normal vision and hearing, and having no severe psychological or neurological conditions. All participants provided written informed consent and received $20 per hour for their participation. Two additional individuals participated, but their data were not recorded (N = 1) or the data timestamp procedure failed (N = 1). The study was approved by the Research Ethics Board at Baycrest Academy for Research and Education.

2.2 Experiment Setup and Procedure

Participants sat in a sound-attenuating booth with their head stabilized on a chin and forehead rest, positioned approximately 70 cm from a computer monitor. Auditory stimuli were presented binaurally through Sony MDR-7506 headphones, with sound played through a Steinberg UR22 mkII external sound card. The experiment was programmed in MATLAB using Psychtoolbox (v3.0.14). The main task was a speech-in-noise perception task. Participants listened to 180 Harvard sentences (mean duration = 2.5 s; Institute of Electrical and Electronics Engineers, 1969) presented in a 7-s 12-talker babble noise (Bilger, 1984), either under an easy (+11 dB SNR) or difficult (-4 dB SNR) listening condition. The sentence started 1.5-s after babble onset and the SNR was determined by adjusting the level of the speech, whereas the level of the babble was constant across trials (Kadem et al., 2020; Ohlenforst et al., 2017). The computer screen remained blank before and during sound presentation. After each sentence, a probe word appeared on the computer screen, and participants judged whether it was semantically related to the preceding sentence (Cui & Herrmann, 2023; Kadem et al., 2020). Participants completed 5 blocks (although 3 participants missed 1, 2, or 3 blocks of data due to incidental technical issues). Each block

comprised 36 trials (18 per SNR condition), and trials were pseudo-randomized such that a maximum of 2 trials of the same SNR could occur in direct succession. A 10-trial practice block preceded the experiment. Behavioral performance in the semantic-relatedness task was calculated as the proportion of correct responses. A paired-samples t-test was used to assess the difference between easy and difficult SNRs.

2.3 Recording of Stationary and Mobile Eye-Trackers

To enable a direct comparison of the two eye-tracking systems, we recorded participants' eye movements and pupil size concurrently using both a stationary and a mobile eye-tracker. An Eyelink DUO was used as a high-precision stationary eye-tracker to record eye movements and pupil size for both eyes at a sampling rate of 1000 Hz. We refer to this device as the "stationary" eye tracker hereafter. A standard nine-point calibration and validation procedure (McIntire et al., 2014) was performed at the beginning of each experimental block. A timestamp message was sent to the eye tracker at the beginning of each babble onset. We used Pupil Labs Neon mobile eye-tracking glasses with a Motorola Edge 40 Pro Android smartphone running the Neon Companion app (v2.8.2) (Baumann & Dierkes, 2023) for data collection and management. We refer to this device as the "mobile" eye tracker hereafter. Eye-movement and pupil-size data were recorded at 200 Hz, whereas head movement data were collected at 110 Hz. The calibration for the mobile glasses involved two steps. First, each participant's intraocular distance was measured and entered into the recording app. Second, participants completed a brief, single-point calibration by fixating on a designated target. On the laptop that was used for stimulus presentation, we also used SocialEyes software written in Python (Saxena et al., 2025) to remotely control the smart phone and Neon Companion app (e.g., start and stop recording). We used SocialEyes' timestamp synchronization procedure to estimate the offset between the internal clocks of the stimulus computer and the recording smartphone (Saxena et al., 2025). Offsets between the two devices clocks were estimated approximately every 10-s during data collection for accurate timestamp calculation during offline analysis.

2.4 Eye-movement, pupil-size, and head-movement data processing

Data were processed offline using MATLAB and Python scripts. The x-coordinate, y-coordinate, and pupil-size data from the stationary eye-tracker were down-sampled to 200 Hz to match the sampling frequency of the eye data from the mobile eye-tracker. Mobile eye-tracking data were corrected for lens distortion using built-in functions from the OpenCV library (Bradski, 2000). Timestamps were estimated through linear interpolation using the recorded offsets between the independent internal clocks of the stimulus computer and the recording smartphone (Saxena et al., 2025). The remaining steps for data processing were similar for both eye trackers.

Pupil-size and x- and y- coordinate data were set to *NaN* (not a number in MATLAB) for 100-ms before and 200-ms after each blink, as well as any outliers that deviated by more than 3 standard deviations from the participant's block mean. All removed and missing data points (*NaN*-coded) were interpolated using a shape-preserving piecewise cubic interpolation (*pchip*). Pupil-size data was then low-pass filtered at 5 Hz (FIR, 131 points, Kaiser Window).

We calculated gaze dispersion as a broad measure of eye movements. We recently developed this measure to non-specifically capture any eye-movement changes associated with listening effort (Cui & Herrmann, 2023; Herrmann et al., 2025; Herrmann & Ryan, 2024). Gaze dispersion was calculated as the log-transformed standard deviation of gaze coordinates (averaged across x and y) within a 1-s sliding window (ignoring data points that had been NaN-coded during preprocessing; and requiring at least 20 data points within a 1-s time window). A smaller value indicates fewer eye movements.

The continuous pupil-size and gaze-dispersion data were divided into epochs ranging from -1 s to 7 s time-locked to babble onset. The percentage of data points within the -1 to 7 epoch that were marked as NaN during pre-processing did not differ between the stationary device (M = 21.49%, SD = 19.24%) and the mobile device (M = 21.64%, SD = 13.57%), $t(25) = 0.029$, $p = 0.977$. Epochs containing equal to or over 50% of samples that were marked as NaN during pre-processing were excluded from further analysis (stationary: M = 13.27, SD = 18.57%; mobile: M = 9.58%, SD = 16.29%; $t(25) = 0.677$, $p = 0.505$).

Epochs were averaged separately for easy and difficult SNRs. To enable cross-device comparisons (stationary vs mobile), both pupil-size and gaze-dispersion time courses (using across-trial average) were converted to within-participant z-scores, computed separately for the stationary and mobile devices to account for scaling differences between devices. The pupil-size time courses were then baseline-corrected by subtracting the mean in the pre-babble period [-1, 0] from all data points, separately for each SNR condition (Mathôt et al., 2018; Winn et al., 2018b). Gaze-dispersion data do not require baseline correction.

Gyroscope data from the mobile glasses were also analyzed to assess SNR effects on head movements (gyroscope data are not available from stationary eye trackers). To this end, head-movement data were analyzed similarly to gaze dispersion as positional variability, calculating the log-transformed standard deviation of the gyroscope values (yaw, pitch, and roll) within a 1-s sliding time window. A lower value indicates fewer head movements. Positional-variability data were divided into epochs ranging from -1 s to 7 s time-locked to babble onset, and averaged across trials for each SNR condition.

For statistical analyses, gaze dispersion and positional variability were averaged within the 1.5 s to 4 s (1.5 s corresponding to sentence onset; 4 s corresponding to the average sentence offset), separately for the easy and difficult listening conditions. For the pupil size, data were averaged within the 2 s to 4.5 s to account for the known delay in the pupil response (Knapen et al., 2016; Winn et al., 2018a, 2018b; Zhang et al., 2022). Both pupil dilation and gaze dispersion were analyzed using a repeated-measures analysis of variance (rmANOVA). This analysis included two factors: SNR (listening condition: easy vs. difficult) and Eye Tracker (stationary vs. mobile). Pearson correlations were used to evaluate the consistency between the two data from the two eye-tracking devices. A paired-sample t-test was used to assess SNR effect on positional variability (head movements). Effect sizes for ANOVA are reported as omega squared (ω^2), and for t-tests, as Cohen's d (d). All analyses were conducted using MATLAB and JASP software (version 0.95.1). Note that data from only 24 participants were available for the analysis of the pupil size. Pupil-size data from two participants were not successfully uploaded to the Pupil Cloud (https://cloud.pupil-labs.com; where initial Pupil Labs processing is calculated), and could also not be recovered offline.

3. Results

3.1 Behavioral Performance

As expected, behavioral performance accuracy was lower for the difficult (M = 0.77, SD = 0.08) than the easy SNR condition (M = 0.93, SD = 0.05; $t(25) = 14.30$, $p = 2.78 \cdot 10^{-13}$, $d = 2.8$).

3.2 Pupil Size

The pupil size was larger in the difficult than in the easy condition (SNR effect: $F(1, 23) = 10.12$, $p = 0.004$, $\omega^2 = 0.12$; Figure 1B). The main effect of Eye Tracker was not significant, $F(1, 23) = 3.66$, $p = 0.068$, nor was the interaction, $F(1, 23) = 1.46$, $p = 0.238$, suggesting that the SNR effect on pupil size did not differ between eye-tracking devices (see also Figure 1C). We confirmed that the SNR effect was indeed significant for both devices (stationary: $t(23) = 4.33$, $p = 2.5 \cdot 10^{-4}$, $d = 0.88$; mobile: $t(23) = 2.285$, $p = 0.032$, $d = 0.47$). Pearson correlations indicated strong associations between the two eye trackers in both the easy ($r = 0.74$, $p < 0.001$) and difficult ($r = 0.88$, $p < 0.001$) listening conditions, suggesting high consistency in pupil size measurements across devices (see Figure 1D).

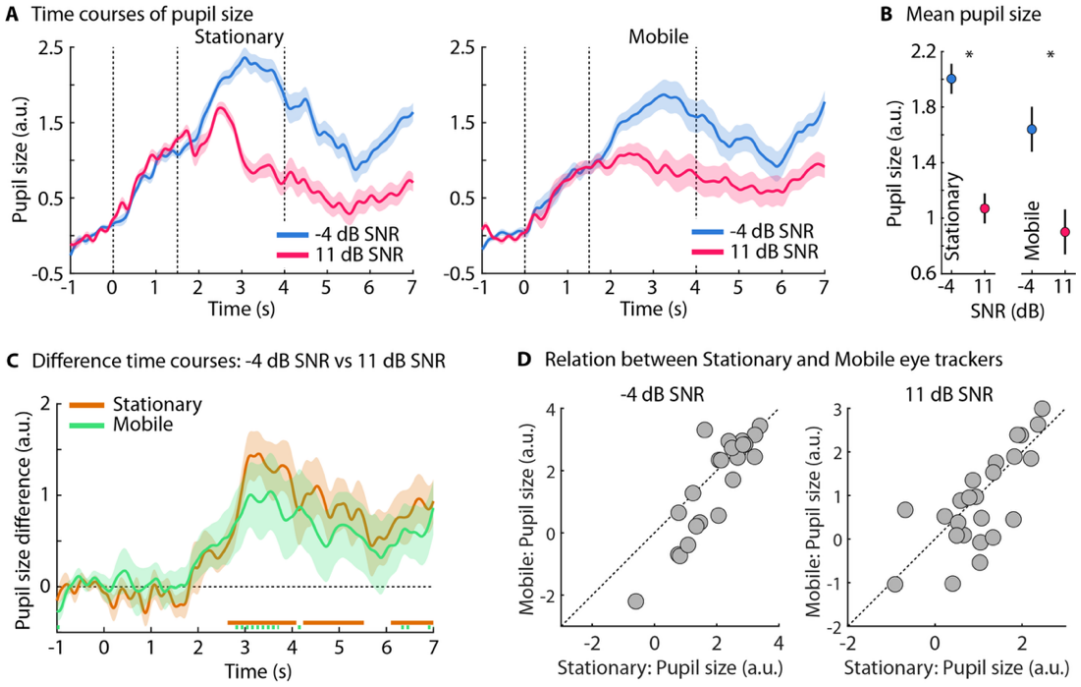

Fig. 1. Pupil responses to listening conditions for stationary and mobile eye trackers. (A) Time courses of pupil size (in z-score (within participant); arbitrary units; a.u.) for stationary (left) and mobile (right) eye trackers. Data are presented for difficult (-4 dB SNR; blue) and easy (+11 dB SNR; pink) listening conditions, with the corresponding shaded areas representing standard error of the mean (between-participant variance removed; (Masson et al., 2003)). Dashed lines denote babble onset (0 s), the sentence onset (1.5 s) and average sentence offset (4 s). (B) Mean pupil size (averaged across 2-4.5 s) for each condition and device. Asterisks denote statistical significance ($p < .05$). (C) Time course of the difference in pupil size between SNR conditions for the stationary (orange) and mobile (green) eye trackers. Horizontal lines at the bottom indicate time points of a statistically significant difference from zero (i.e., significant SNR effect; $p < .05$; solid line = corrected using False Discovery Rate (Benjamini & Hochberg, 1995); dashed line = uncorrected). (D) Scatter plots show the correlation between stationary and mobile eye trackers for pupil size during difficult (left) and easy (right) conditions. Data points above the diagonal line indicate that the mobile eye tracker measured a larger pupil size, while points below the line show a larger pupil size measured by the stationary eye tracker.

3.3 Gaze Dispersion

Gaze dispersion was lower for the difficult compared to the easy listening condition (SNR effect: $F(1,25) = 118.77$, $p = 2.92 \cdot 10^{-11}$, $\omega^2 = 0.52$; Figure 2). The effect of Eye Tracker was not significant ($F(1,25) = 0.28$, $p = 0.605$), nor the interaction ($F(1,25) = 1.30$, $p = 0.265$), suggesting gaze dispersion is similarly sensitive to SNR for both device types. We confirmed that the gaze-dispersion effect was significant for the stationary ($t(25) = 9.89$, $p = 1.47 \cdot 10^{-10}$, $d = 1.94$) and the mobile eye tracker ($t(25) = 7.74$, $p = 1.01 \cdot 10^{-7}$, $d = 1.52$; Figure 2C). Pearson correlations indicated moderate-to-strong correlations between the two eye trackers for both the easy ($r = 0.58$, $p = 0.003$) and difficult SNR ($r = 0.73$, $p = 2.6 \cdot 10^{-6}$; Figure 2D).

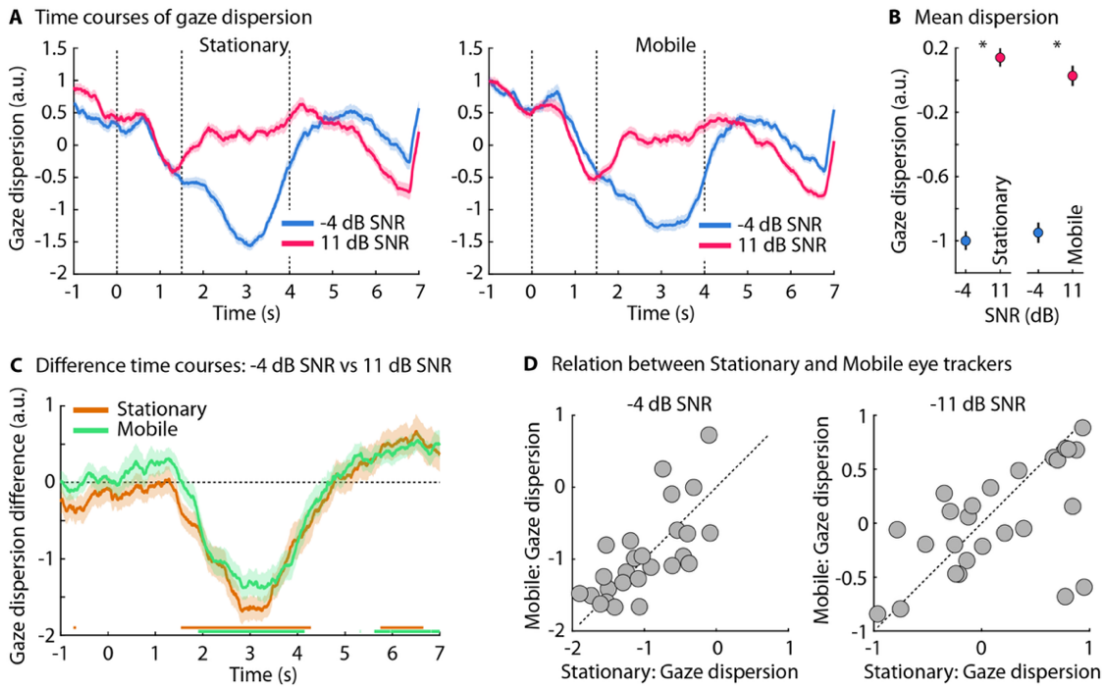

Fig. 2. Gaze dispersion responses to listening conditions for stationary and mobile eye trackers. (A) Time courses of gaze dispersion (in z□score (within participant); arbitrary units; a.u.) for stationary (left) and mobile (right) eye trackers. Data are presented for difficult (-4 dB SNR; blue) and easy (+11 dB SNR; pink) listening conditions, with the corresponding shaded areas representing standard error of the mean (between-participant variance removed; (Masson et al., 2003)). Dashed lines denote babble onset (0 s), the sentence onset (1.5 s) and average sentence offset (4 s). (B) Mean gaze dispersion (averaged across 1.5-4 s) for each condition and device. Asterisks denote statistical significance ($p < .05$). (C) Time course of the difference in gaze dispersion between two conditions for the stationary (orange) and mobile (green) eye trackers. Horizontal lines at the bottom indicate time points of a statistically significant difference from zero (i.e., a significant SNR effect; $p < .05$ corrected using False Discovery Rate (Benjamini & Hochberg, 1995)). (D) Scatter plots show the correlation between stationary and mobile eye trackers for gaze dispersion during difficult (left) and easy (right) conditions. Data points above the diagonal line indicate that the mobile eye tracker measured a larger gaze dispersion, while points below the line show a larger gaze dispersion measured by the stationary eye tracker.

3.4 Head Movements

Head movements – that is, positional variability – measured using the mobile eye-tracking glasses was lower for the difficult compared to the easy SNR ($t(25) = 2.97$, $p = 0.006$, $d = 0.58$; Figure 3).

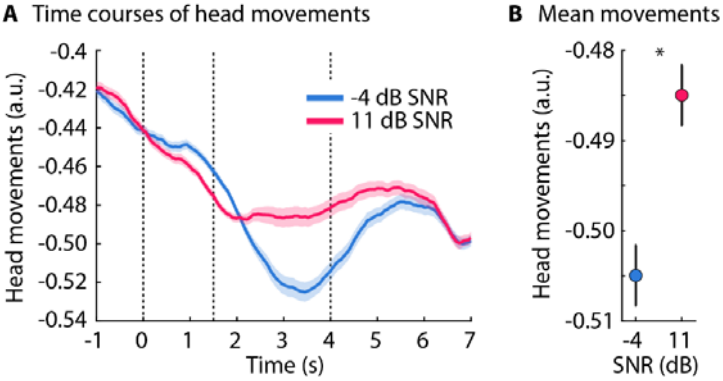

Figure 3. Positional variability of head movements, measured by mobile eye-tracking glasses, index of listening effort. (A) Time courses of positional variability of head movements (in arbitrary units; a.u.). Data are presented for the difficult (-4 dB SNR; blue) and easy (+11 dB SNR; pink) listening conditions. Shaded regions represent the

standard error of the mean (between-participant variance removed; (Masson et al., 2003)). Vertical dashed lines mark the babble onset (0 s), sentence onset (1.5 s) and average sentence offset (4 s). (B) Mean head movements (averaged across 1.5-4 s) for each condition. Lower values on the y-axis indicate more stable head position. *$p < .05$.

3.5 Relationships Between Listening Effort Measures

Correlation analyses were conducted using the difference scores between the difficult and easy conditions for each measure (i.e., SNR effect). There were no significant relationships between the listening effort markers: SNR effects for gaze dispersion and pupil dilation were not correlated for stationary (r = -0.20, p = 0.340, 95 % CI = [-0.56, 0.22]) nor mobile eye-tracking data (r = 0.04, p = 0.860, 95 % CI = [-0.37, 0.44]). Further analyses using the data from the mobile eye-tracking glasses revealed a marginally significant correlation between positional variability (head movement) and gaze dispersion (r = 0.39, p = 0.051, 95 % CI = [-0.016, 0.69]) and a significant negative correlation with pupil dilation (r = -0.44, p = 0.026, 95 % CI = [0.045, 0.72]). Noteworthy, these correlation analyses were post-hoc, therefore, they might be underpowered due to a small sample size (N = 24).

4. Discussion

The current study investigated whether a pair of state-of-the-art mobile eye-tracking glasses are sensitive to several physiological markers of listening effort. We found reduced gaze dispersion, increased pupil size, and reduced head movements when speech masking – and thus listening effort – increased. Cross-device comparisons found no significant differences in the speech-masking effect, although there was some indication that the pupil-size effect may be numerically smaller for the mobile than the stationary eye-tracker. Overall, the current results indicate high feasibility of using mobile eye-tracking glasses for the assessment of listening effort, including head-movement measures that are not available using a stationary eye-tracker.

4.1 Ocular and Head Markers

Both eye-trackers converged on the same pattern: as listening demands increased, gaze dispersion decreased, whereas pupil size increased. These results are highly consistent with previous studies using pupillometry (Almaraz et al., 2018; Piquado et al., 2010; Van Der Wel & Van Steenbergen, 2018; Zekveld et al., 2018) and eye movements (Contadini-Wright et al., 2023; Cui & Herrmann, 2023; Herrmann et al., 2025; Herrmann & Ryan, 2024) to assess listening effort. An increase in pupil size and a reduction in eye movements appear to capture complementary aspects of listening effort (see also (Liu & Chait, 2025)). These measures likely tap into interconnected functions within overlapping networks. For example, pupil size is linked to arousal involving the locus coeruleus-norepinephrine systems (Joshi et al., 2016), which is subject to top-down regulation from prefrontal control regions (Joshi & Gold, 2020; Ross & Van Bockstaele, 2021). Meanwhile, gaze dispersion may index effort through attentional control that suppresses visual exploration, possibly relying on involvement of superior colliculus, thalamus, basal ganglia, cerebellum, and the prefrontal cortex (Lencer et al., 2019; Pierce et al., 2019; Schneider et al., 2014; Sparks, 2002).

Despite the use of a chin and forehead rest, participants still showed reduced head movements as the listening task became more challenging. These results show that head movements persist despite the use of a chin and forehead rest, and it still could be modulated by task demands. Furthermore, the reduction in head movements during effortful listening suggests that individuals may have adopted a freezing or stabilization strategy to free up cognitive resources. By reducing head movements, this strategy could minimize the amount of proprioceptive information (Dault et al., 2001), thereby freeing up more cognitive resources for processing acoustic information in the speech-in-noise task (Assländer & Peterka, 2014). This principle is further supported by neurophysiological evidence from rodents, which shows that reduced body movement is linked to heightened neuronal activity in the auditory cortex (Schneider et al., 2014). Critically, although we observed a marginal correlation between head movements and eye-tracking measures (pupil-size and gaze dispersion), it is unlikely that the SNR effects on gaze dispersion and pupil

size are simple artifacts of head movements. For example, the mobile eye-tracking glasses are head-mounted, and the recorded eye movements should thus be unaffected by head movements. Nevertheless, the relationship between head movement and ocular changes presents an interesting area for future research.

4.2 Device Comparison

The effect of SNR (speech masking) on gaze dispersion was comparable across the stationary and the mobile eye tracker. For pupil size, the SNR effect was numerically smaller for the mobile compared to the stationary eye-tracker. However, neither the Device main effect nor the Device × SNR interaction was significant, thus providing no formal evidence for a difference between devices. Without over-interpreting the numerically smaller SNR effect for the mobile glasses, potential contributors could be the mobile system's less precise calibration, stationary geometry vs head-worn geometry for estimating the pupil size (Brisson et al., 2013; Hayes & Petrov, 2016), or generally a lower sensitivity. Calibration for the mobile glasses was carried out once at the beginning of the experiment, whereas the EyeLink eye tracker was calibrated prior to each block. Moreover, calibration quality in head-worn systems could also degrade more rapidly over time, at least for older Pupil Labs models and software (Ehinger et al., 2019).

Critically, in addition to finding no significant differences for the SNR effect between the EyeLink and the Pupil Labs eye-trackers, individual participant data correlated between the two devices. This suggests that state-of-the art mobile eye-tracking glasses, such as the Pupil Labs Neon, provide unique opportunities to replace more expensive, stationary eye-trackers with mobile eye-tracking glasses without much loss in sensitivity to assess listening effort. In fact, mobile eye-tracking glasses provide additional benefits by measuring head movements, which also are indicative of listening effort.

4.3 Conclusion

The current study compared ocular markers of listening effort between a standard stationary and a modern mobile eye-tracker. Both the stationary and the mobile eye-trackers showed increased pupil sizes and reduced eye movements as listening became effortful, with no significant differences between the devices. Mobile eye-tracking glasses additionally revealed reduced head movements under higher listening demand, which is not available using stationary eye-tracking systems. The current study indicates that mobile eye-tracking reliably captures listening effort, opening a path towards the assessments of listening effort in real life; however, as the present data were collected under constrained laboratory conditions, further validation in more naturalistic environments is warranted.

Acknowledgments

This work was supported by grants from the Canada Research Chair program (CRC-2023-00383; BH), the Natural Sciences and Engineering Research Council of Canada (RGPIN-2021-02602 to BH; RGPIN-2023-05050 to LKF), and the Canadian Institutes of Health Research (186236; BH).

Author Declarations

Conflict of Interest:

The authors have no conflicts of interest to disclose.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Alemu, R. Z., Papsin, B. C., Harrison, R. V., Blakeman, A., & Gordon, K. A. (2024). Head and Eye Movements Reveal Compensatory Strategies for Acute Binaural Deficits During Sound Localization. *Trends in Hearing*, *28*. https://doi.org/10.1177/23312165231217910/

Almaraz, D., Carlson, B., Grafelman, E., Ingalls, J., Quintanilla, Y., Vu, H.-T., & Loebach, J. (2018). Pupillometry as a measure of auditory cognitive processes and listening effort. *The Journal of the Acoustical Society of America*, *143*(3_Supplement), 1751–1751. https://doi.org/10.1121/1.5035727Arachchige, S. N., Burch V, R. F., Chander, H., Turner, A. J., & Knight, A. C. (2022). The use of wearable devices in cognitive fatigue: current trends and future intentions. Theoretical Issues in Ergonomics Science, 23(3), 374-386.

Baumann, C., & Dierkes, K. (n.d.). *Neon Accuracy Test Report*. https://doi.org/10.5281/ZENODO.10420388

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/J.2517-6161.1995.TB02031.X

Bradski, G. (2000) The OpenCV Library. Dr. Dobb's Journal of Software Tools, 120; 122-125. - References - Scientific Research Publishing. (n.d.). Retrieved September 8, 2025, from https://www.scirp.org/reference/ReferencesPapers?ReferenceID=1692176

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. *Behavior Research Methods*, *45*(4), 1322–1331. https://doi.org/10.3758/S13428-013-0327-0/

Chavant, M., & Kapoula, Z. (2022). Eye-Movement Deficits in Seniors with Hearing Aids: Cognitive and Multisensory Implications. *Brain Sciences 2022, Vol. 12, Page 1425*, *12*(11), 1425. https://doi.org/10.3390/BRAINSCI12111425

Contadini-Wright, C., Magami, K., Mehta, N., & Chait, M. (2023). *Behavioral/Cognitive Pupil Dilation and Microsaccades Provide Complementary Insights into the Dynamics of Arousal and Instantaneous Attention during Effortful Listening*. https://doi.org/10.1523/JNEUROSCI.0242-23.2023

Craighero, L. (2022). The role of the sensorimotor system in cognitive functions. *Brain sciences*, *12*(5), 604.

Cui, M. E., & Herrmann, B. (2023). *Behavioral/Cognitive Eye Movements Decrease during Effortful Speech Listening*. https://doi.org/10.1523/JNEUROSCI.0240-23.2023

Dault, M. C., Geurts, A. C., Mulder, T. W., & Duysens, J. (2001). Postural control and cognitive task performance in healthy participants while balancing on different support-surface configurations. *Gait & posture*, *14*(3), 248-255.Ehinger, B. V., Groß, K., Ibs, I., & König, P. (2019). A new comprehensive eye-tracking test battery concurrently evaluating the Pupil Labs glasses and the EyeLink 1000. *PeerJ*, *2019*(7), e7086. https://doi.org/10.7717/PEERJ.7086/SUPP-1

Fink, L., Simola, J., Tavano, A., Lange, E., Wallot, S., & Laeng, B. (2024). From pre-processing to advanced dynamic modeling of pupil data. *Behavior Research Methods*, *56*(3), 1376–1412. https://doi.org/10.3758/S13428-023-02098-1/

H., R. E. (1969). IEEE Recommended Practice for Speech Quality Measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*(3), 225–246. https://doi.org/10.1109/TAU.1969.1162058

Hayes, T. R., & Petrov, A. A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods*, *48*(2), 510–527. https://doi.org/10.3758/S13428-015-0588-X/

Helfer, K. S., & Jesse, A. (2021). Hearing and speech processing in midlife. *Hearing Research*, *402*, 108097. https://doi.org/10.1016/J.HEARES.2020.108097

Herrmann, B., & Johnsrude, I. S. (2020). A model of listening engagement (MoLE). *Hearing Research*, *397*, 108016. https://doi.org/10.1016/J.HEARES.2020.108016

Herrmann, B., & Ryan, J. D. (2024). Pupil Size and Eye Movements Differently Index Effort in Both Younger and Older Adults. *Journal of Cognitive Neuroscience*, *36*(7), 1325–1340. https://doi.org/10.1162/JOCN_A_02172

Herrmann, B., Scharf, F., & Widmann, A. (2025). Eye movements of younger and older adults decrease during story listening in background noise. *BioRxiv*, 2025.04.15.648919. https://doi.org/10.1101/2025.04.15.648919

Hidalgo, C., Zielinski, C., Chen, S., Roman, S., Truy, E., & Schön, D. (2023). *Perception of Speech Turn Dynamics is preserved in Congenitally Deaf children with Cochlear Implants.* https://doi.org/10.1101/2023.05.22.538554

Higgins, N. C., Pupo, D. A., Ozmeral, E. J., & Eddins, D. A. (2023). Head movement and its relation to hearing. *Frontiers in Psychology*, *14*, 1183303. https://doi.org/10.3389/FPSYG.2023.1183303/XML

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. https://doi.org/10.1016/J.NEURON.2015.11.028/ASSET/076AE6C2-4035-40ED-8AF1-7B3D95E3BC16/

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

Kadem, M., Herrmann, B., Rodd, J. M., & Johnsrude, I. S. (2020). Pupil Dilation Is Sensitive to Semantic Ambiguity and Acoustic Degradation. *Trends in Hearing*, *24*. https://doi.org/10.1177/2331216520964068/ASSET/580923B7-B311-44D7-9F01-732AB3D1FA4C/

Knapen, T., De Gee, J. W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PLoS ONE*, *11*(5). https://doi.org/10.1371/JOURNAL.PONE.0155574

Lencer, R., Sprenger, A., & Trillenberg, P. (2019). *Smooth Eye Movements in Humans: Smooth Pursuit, Optokinetic Nystagmus and Vestibular Ocular Reflex*. 117–163. https://doi.org/10.1007/978-3-030-20085-5_4

Liu, X., & Chait, M. (2025). Dissociable Pupil and Oculomotor Markers of Attention Allocation and Distractor Suppression during listening. bioRxiv, 2025-05.

Masson, M. E. J., Loftus, G. R., & Masson, M. (2003). Using Confidence Intervals for Graphically Based Data Interpretation. *Canadian Journal of Experimental Psychology*.

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, *50*(1), 94–106. https://doi.org/10.3758/S13428-017-1007-2/

McIntire, L. K., McIntire, J. P., McKinley, R. A., & Goodyear, C. (2014). Detection of vigilance performance with pupillometry. *Eye Tracking Research and Applications Symposium (ETRA)*, 167–174. https://doi.org/10.1145/2578153.2578177

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., & 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. https://doi.org/10.1016/J.HEARES.2017.05.012

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Perea Perez, F. (2023). *Real-world listening effort in adult cochlear implant users.*

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016a). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016b). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Pierce, J. E., Clementz, B. A., & McDowell, J. E. (2019). *Saccades: Fundamentals and Neural Mechanisms*. 11–71. https://doi.org/10.1007/978-3-030-20085-5_2

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560–569. https://doi.org/10.1111/J.1469-8986.2009.00947.XRoss, J. A., & Van Bockstaele, E. J. (2021). The locus coeruleus-norepinephrine system in stress and arousal: Unraveling historical, current, and future perspectives. Frontiers in Psychiatry, 11, 601519. https://doi.org/10.3389/fpsyt.2020.601519, PubMed: 601519, PubMed: 33584368

Saxena, S., Visram, A., Lobo, N., Mirza, Z., Khan, M., Pirabaharan, B., Nguyen, A., & Fink, L. K. (2025). SocialEyes: Scaling Mobile Eye-tracking to Multi-person Social Settings. *Conference on Human Factors in Computing Systems - Proceedings* . https://doi.org/10.1145/3706598.3713910/

Schneider, D. M., Nelson, A., & Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature 2014 513:7517*, *513*(7517), 189–194. https://doi.org/10.1038/nature13724

Sparks, D. L. (2002). The brainstem control of saccadic eye movements. *Nature Reviews Neuroscience 2002 3:12*, *3*(12), 952–964. https://doi.org/10.1038/nrn986

Thurman, S. M., Cohen Hoffing, R. A., Madison, A., Ries, A. J., Gordon, S. M., & Touryan, J. (2021). "Blue Sky Effect": Contextual Influences on Pupil Size During Naturalistic Visual Search. *Frontiers in Psychology*, *12*, 748539. https://doi.org/10.3389/

Van Der Wel, P., & Van Steenbergen, H. (2018). *Pupil dilation as an index of effort in cognitive control tasks: A review*. https://doi.org/10.3758/s13423-018-1432-y

Vandemoortele, S., Feyaerts, K., Reybrouck, M., De Bièvre, G., Brône, G., & De Baets, T. (2018). Gazing at the partner in musical trios: a mobile eye-tracking study. *Journal of Eye Movement Research*, *11*(2), 10.16910/jemr.11.2.6. https://doi.org/10.16910/JEMR.11.2.6

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018a). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends in Hearing*, *22*. https://doi.org/10.1177/2331216518800869/ASSET/01626141-1861-4DB8-BE62-C110C0999A39/

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018b). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends in Hearing*, *22*. https://doi.org/10.1177/2331216518800869/

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in Hearing*, *22*. https://doi.org/10.1177/2331216518777174/ASSET/2150B2DC-C9EA-43A6-9B09-FF05FACA5B6E/

Zhang, Y., Malaval, F., Lehmann, A., & Deroche, M. L. D. (2022). Luminance effects on pupil dilation in speech-in-noise recognition. *PLOS ONE*, *17*(12), e0278506. https://doi.org/10.1371/JOURNAL.PONE.0278506