Seeing an Auditory Object: Pupillary Light Response Reflects Covert Attention to Auditory Space and Object

Liao, Hsin-I
NTT Communication Science Laboratories

Fujihira, Haruna
NTT Communication Science Laboratories

Yamagishi, Shimpei
NTT Communication Science Laboratories

Yang, Yung-Hao
NTT Communication Science Laboratories

他

出版情報：Journal of Cognitive Neuroscience. 35 (2), pp.276-290, 2023-02-01. MIT Press
バージョン：

Seeing an Auditory Object: Pupillary Light Response Reflects Covert Attention to Auditory Space and Object

Hsin-I Liao[1], Haruna Fujihira[1,2], Shimpei Yamagishi[1], Yung-Hao Yang[1], and Shigeto Furukawa[1]

Abstract

■ Attention to the relevant object and space is the brain's strategy to effectively process the information of interest in complex environments with limited neural resources. Numerous studies have documented how attention is allocated in the visual domain, whereas the nature of attention in the auditory domain has been much less explored. Here, we show that the pupillary light response can serve as a physiological index of auditory attentional shift and can be used to probe the relationship between space-based and object-based attention as well. Experiments demonstrated that the pupillary response corresponds to the luminance condition where the attended auditory object (e.g., spoken sentence) was located, regardless of whether attention was directed by a spatial (left or right) or nonspatial (e.g., the gender of the talker) cue and regardless of whether the sound was presented via headphones or loudspeakers. These effects on the pupillary light response could not be accounted for as a consequence of small (although observable) biases in gaze position drifting. The overall results imply a unified audiovisual representation of spatial attention. Auditory object-based attention contains the space representation of the attended auditory object, even when the object is oriented without explicit spatial guidance. ■

INTRODUCTION

The auditory world is rarely silent. It is usually full of various sounds, including background noises, and sometimes with multiple people talking at the same time. Attention to the relevant object and space is the brain's strategy to effectively process the information of interest. A massive number of studies based on behavioral measurements have described how attention is distributed in visual space, for example, as the spotlighting metaphor (Posner, Snyder, & Davidson, 1980), the zoom-lens metaphor (Cave & Bichot, 1999; Eriksen & St. James, 1986), or the gradient model (Downing, 1988). However, the evidence for the nature of "auditory" spatial attention is less robust and controversial (Best, Shinn-Cunningham, Ozmeral, & Kopco, 2010; Best, Ozmeral, Kopčo, & Shinn-Cunningham, 2008; Spence & Driver, 1994, 1996; Mondor & Zatorre, 1995; Quinlan & Bailey, 1995; Rhodes, 1987). This makes it difficult to infer how spatial attention functions in auditory space and for an auditory object. Neuroimaging studies indicate that auditory spatial attention and visual spatial attention share the same neural circuit underlying the dorsal frontoparietal cortical networks (Braga, Fu, Seemungal, Wise, & Leech, 2016; Smith et al., 2010; Corbetta, 1998). It suggests that auditory spatial attention operates in a way similar to visual spatial attention. In the current study, we aimed to demonstrate that the internal processes related to auditory attention can be "read out" via the pupillary light response (PLR) as they have been shown to be in visual attention (Mathôt, van der Linden, Grainger, & Vitu, 2013). Moreover, PLR tracks auditory spatial attention across time, presumably reflecting the internal processing of how attention selects an auditory object.

Recent evidence indicates that the PLR is modulated by neural activities from the FEF (Ebitz & Moore, 2017), and this neural basis may explain why it reflects not only changes in physical luminance but also top–down modulated perceptual brightness (also see Strauch, Wang, Einhäuser, Van der Stigchel, & Naber, 2022; Binda & Gamlin, 2017). Pioneering studies have demonstrated that pupils respond to perceived brightness even when the physical luminance input remains the same in the case of visual awareness to luminance in binocular rivalry (Naber, Frässle, & Einhäuser, 2011), the bright illusion (Suzuki, Minami, Laeng, & Nakauchi, 2019; Laeng & Endestad, 2012), visual scene interpretation (Binda, Pereverzeva, & Murray, 2013; Naber & Nakayama, 2013), and mental imagery (Laeng & Sulutvedt, 2014). Remarkably, pupils also reflect the luminance condition of the location to which visual attention is directed while the eyes fixate steadily (Strauch, Romein, Naber, Van der Stigchel, & Ten Brink, 2022; Binda, Pereverzeva, & Murray, 2014; Mathôt, Dalmaijer, Grainger, & Van der Stigchel, 2014; Binda et al., 2013; Mathôt et al., 2013; Haab, 1886). For

[1]NTT Communication Science Laboratories, Japan, [2]Japan Society for the Promotion of Science

Journal of Cognitive Neuroscience 35:2, pp. 276–290
https://doi.org/10.1162/jocn_a_01935

instance, Mathôt et al. (2013) demonstrated that when participants view a visual display containing luminance disparity between the left and right visual hemifields (e.g., darkness on the left and brightness on the right, or vice versa), pupil size is larger when they covertly attend to, but not overtly shift the eyes to, the dark rather than bright visual hemifield. A recent study further showed that this attentional modulation of the PLR operates on abstract mental content of a left-to-right spatial–numerical association (Salvaggio, Andres, Zénon, & Masson, 2022). This implies that the attentional modulation of the PLR does not only serve for the anticipation of the coming perception in the visual system (Mathôt, 2018; Mathôt & Van der Stigchel, 2015) but automatically interacts with a more general cognitive attentional function as well. This view also suggests the possibility of attentional modulation of PLR in other sensory domains.

One apparent difference between visual attention and auditory attention is that visual attention is better understood in its spatial deployment, whereas auditory attention enables us to understand more about its temporal aspect, assuming that both operate through the same underlying neural network (Noyce, Kwasa, & Shinn-Cunningham, 2022). Auditory objects or streams consist of acoustic features extended through time, which can be presented in separate or overlapped space (Shinn-Cunningham, 2008; Fritz, Elhilali, David, & Shamma, 2007). Space is not considered an essential feature to define an auditory object. By contrast, in natural scenes, visual objects are generally associated with particular locations in space. Theoretically, the location serves as the master map in visual attention (Treisman & Gelade, 1980). Assuming the same spatial attention mechanism operates in both the visual and auditory domains, it would be expected that the same attentional modulation of the PLR operates in auditory space, particularly when the attended auditory object is referenced by a spatial cue. However, it is unclear whether the location information is automatically and compulsively represented in the auditory object when it is attended and defined via nonspatial acoustic features. Neuroimaging studies have shown controversial evidence: Some indicate that nonspatial auditory attention engages networks such as the inferior frontal gyrus (Larson & Lee, 2014; Hill & Miller, 2009); others indicate an overlapping neural circuit between space-based and object-based auditory attention (Bushara et al., 1999; Zatorre, Mondor, & Evans, 1999). If nonspatial object-based auditory attention operates independently of space-based auditory attention, it is plausible that the auditory object can be attended and processed without spatial attention allocated to its location. Alternatively, if object-based and space-based auditory attention share a common mechanism, spatial attention is expected to shift to the object's location even when the auditory object is selected via nonspatial features. In addition, the timing of spatial attention shifts may vary if the involvement of spatial information requires further processing depending on the circumstances.

We addressed the above issues by investigating how PLR reflects spatial attention to auditory objects. In four experiments, human participants listened to two concurrent environmental sounds or speech sentences presented dichotically to the two ears through headphones (Experiments 1–3) or through two loudspeakers located left and right in space (Experiment 4). They were instructed to attend the one defined by a spatial cue (left or right) or by a nonspatial cue, for example, the gender of the talker (male or female). Sitting in front of a visual display containing luminance disparity between the left and right visual hemifields, they fixated the center of the display while listening to the auditory stimuli. An infrared video-based eye tracker recorded their pupillary responses and gaze positions throughout the experiments. The results consistently demonstrated that their pupils dilated more strongly when the attended auditory object was located on the dark side of the visual hemifield than when it was on the bright side. The finding was replicated regardless of whether attention was directed by a spatial or nonspatial cue and whether the sound was presented via headphones or loudspeakers. The timing of the PLR divergence (i.e., the difference between attend-to-dark and bright conditions) occurred earlier for the spatial cue than nonspatial cue. Local luminance differences due to gaze position drift could not explain most of the effects. The overall results imply that auditory attention to an object in space and visual spatial attention recruit a common underlying mechanism. When the auditory object is directed by the nonspatial cue, extra time is needed to identify the auditory object's location before shifting spatial attention accordingly. The finding provides profound insights into not only the neural mechanism of spatial attention across modalities but also into brain–computer interface to predict human auditory spatial attention by eyes.

METHODS

Participants

Seventy-four adults (54 women, age range 20–50 years, median age 39 years) who had normal or corrected-to-normal vision and normal hearing acuity participated in the current study (15 in Experiment 1, 15 in Experiment 2, 27 in Experiment 3, and 17 in Experiment 4). Sample sizes were chosen based on our previous studies with comparable pupillometry measurements and trial numbers per participant (Liao, Kashino, & Shimojo, 2021; Liao, Kidani, Yoneya, Kashino, & Furukawa, 2016). In Experiment 3, the number of trials in the critical dichotic condition for each participant was half of that in Experiment 2, because of adding a diotic sound presentation condition (see Design and Procedure for details). To have a similar number of critical trials, more participants were recruited. One participant in Experiment 4 was excluded because of the < 50% accuracy in task performance (46.9%). All were naive about the purpose of the study. The current

study was approved by the NTT Communication Science Laboratories ethics committee. All participants gave written informed consent before the experiment and received payment for their participation.

Overview of Experiments

Experiment 1 was modified from Mathôt et al. (2013) by changing the main task to an auditory task to discriminate two environmental sounds. Experiments 2, 3, and 4 used spoken sentences. Participants dichotically listened to two sentences spoken by different talkers, one male and one female, and attended the one according to the instruction. In Experiment 2, the target sentence was defined by space and the gender of the talker. In Experiments 3 and 4, the target sentence was defined by gender. In Experiment 3, half of the target sentences were presented dichotically, and the other half was presented diotically. In Experiment 4, the auditory stimuli were presented via two loudspeakers instead of the headphones as in Experiments 1–3.

All the experiments had a general common structure as follows. A trial started with presentation of a visual display with its left and right hemifields dark and bright, respectively, or vice versa. After 3 sec as an adaptation period, a "voice cue" was presented to indicate the target dimension (i.e., space or gender) in Japanese, followed by a "silent period" (Experiment 1) or a "listening period" (Experiments 2–4). This was followed by a "response period" wherein participants were asked to discriminate the environmental sound as soon and accurately as possible (Experiment 1) or to recall the content of the target sentence with no time pressure (the others). Participants were given a written and oral explanation about the nature of the auditory task and performed several practice trials to familiarize themselves with the task.

Apparatus and Stimuli

Auditory and visual stimuli were generated with MATLAB (The MathWorks, Inc.) and PsychToolbox (Kleiner, Brainard, & Pelli, 2007; Brainard, 1997; Pelli, 1997), controlled by a personal computer (Dell OptiPlex 755). Visual stimuli were presented on an 18.1-in. monitor (Eizo FlexScan L685Ex) with a 60-Hz frame rate and a 1280 × 1024 resolution in Experiments 1, 2, and 3 and a 1024 × 768 resolution in Experiment 4. There were two types of visual displays with luminance disparity between the left and right visual fields: black (0.35 cd/m^2) on the left and white (94.04 cd/m^2) on the right, and vice versa. In Experiment 1, a gray fixation cross (0.5° × 0.5°, 21.04 cd/m^2) was presented at the center of the visual display against the background with luminance disparity. In Experiments 2, 3, and 4, a gray vertical band (2.3° in width, 21.04 cd/m^2) acted as a "buffer zone" that was superimposed at the center against the background with luminance disparity to lessen the sharp luminance variation in fixation. A Gabor patch (0.5° × 0.5°), which was generated by superimposing a Gaussian and a sine-wave function with a vertical orientation (10 cycles per degree) and presented at the center, served as the fixation point.

Auditory stimuli were presented through headphones (Sennheiser HD 595) in Experiments 1, 2, and 3 and through two custom-built loudspeakers in Experiment 4. The loudspeakers were positioned so that their centers were located 70 cm left and right of the fixation point on the visual display, and the surfaces of the display and the loudspeakers were aligned. The voice cues were recorded by a female native-Japanese speaker in our laboratory. They had a duration of 400–570 msec and were presented diotically (identical signal to both ears). The target sounds used in Experiment 1, a dog barking and a phone ringing, were sampled from a database on a compact disc (Audio Pro Sound Effects by Yannick Chevalier) and edited to be 500 msec in length. In Experiments 2, 3, and 4, the target sentences were sampled from the coordinate response measure (CRM) corpus recorded in our laboratory, Japanese version with a slight variation in the keywords of Bolia, Nelson, Ericson, and Simpson (2000). Each CRM sentence consisted of three Japanese keywords, namely, an animal name (KUMA [bear], SHIKA [deer], TORA [tiger], INU [dog], NEKO [cat], TORI [bird], SARU [monkey], or BUTA [pig]), a color (AKA [red], AO [blue], SHIRO [white], KURO [black], MOMO [pink], or NIJI [rainbow]), and a number (ZERO [zero], ICHI [one], SAN [three], YON [four], ROKU [six], NANA [seven], HACHI [eight], KYUU [nine], or JUU [ten]). The sentences were spoken by 10 different native-Japanese talkers (five of each gender) in a 2.5- to 3-sec length. The sounds were presented at a comfortable listening level, self-adjusted by each participant.

During the response period, participants viewed the fixation display with luminance disparity in Experiment 1. In Experiments 2, 3, and 4, there were three types of response displays, corresponding to the three keywords, namely, an animal name, color, and number. The animal-response display consisted of eight patches (2.5° × 1.3°), each labeled with animal names. The color-response display consisted of six patches, each labeled with color names. The number-response display consisted of nine patches, each labeled with a number. All the labels were written in Japanese Katakana.

Eye movements including pupillary responses were recorded binocularly with an infrared eye-tracker camera (Eyelink 1000 Plus Desktop Mount, SR Research Ltd.). The camera was positioned below the monitor. The sampling rate of the recording was 1000 Hz. Participants sat in front of the monitor at an 80-cm distance with their head supported on a chin rest. Before each formal experimental session, they went through the 5-point Eyelink calibration program to calibrate and validate their eye data. After the calibration, they were instructed to fixate the central fixation point while performing the auditory task.

Design and Procedure

In Experiment 1 (Figure 1A), the voice cue indicating "left" or "right" was presented, followed by a 5-sec silent period. In the response period, the target sounds (i.e., dog barking and phone ringing) were presented dichotically (one to each ear). Participants were asked to shift their attention to the left or right ear, according to the voice cue's instruction, and discriminate whether the sound presented in the cued direction was "dog barking" or "phone ringing" by pressing a designated key on the keyboard as soon and accurately as possible. The correspondence of the response key and target sound was counterbalanced across the participants: Half of the participants pressed the "M" key for "dog barking" and the "N" key for "phone ringing," and the other half did the opposite. Once they pressed the response key, the next trial started. The visual display disparity (black on left or right), the voice cue (left or right), and the target sound (dog barking sound to the left or right ear) were counterbalanced across trials. There were 48 trials in each block. Each block took about 15 min, and the participants performed two blocks with a 15-min break in between.

In Experiments 2, 3, and 4, 1 sec after the voice cue presentation, the listening period started, in which the two target sentences were presented dichotically (Figure 2A). The target sentences were selected so that the contents of the three keywords would not overlap. Each sentence was repeated twice, resulting in a total duration of 6 sec. Participants were asked to pay attention to the target sentence, memorize its content, and recall it during the response period. With no time pressure, they gave the answer by using the mouse to click the patch in the response display, which was randomly selected from the three types of response displays for each trial. Participants did not know in advance the response set (animal, color, or number) for each trial and thus had to memorize the entire content to perform the task.

Experiment 2 consisted of two conditions, namely, the attend-to-location and attend-to-gender conditions. In the attend-to-location condition, the voice cue was on the spatial dimension ("left" or "right"). Participants paid attention to the sentence presented at the cued direction, as in Experiment 1. In the attend-to-gender condition, the voice cue was on the dimension of the talker's gender ("male" or "female"). Participants paid attention to the sentence spoken by the male talker if the voice cue said "male" beforehand or to the one spoken by the female talker if the voice cue said "female" beforehand. The visual display disparity, the voice cue (left or right in the attend-to-location condition; male or female in the attend-to-gender condition), and the target sound (the male's spoken sentence to the left or right ear) were counterbalanced across trials. There were 64 trials in each block. Participants performed the two types of attention

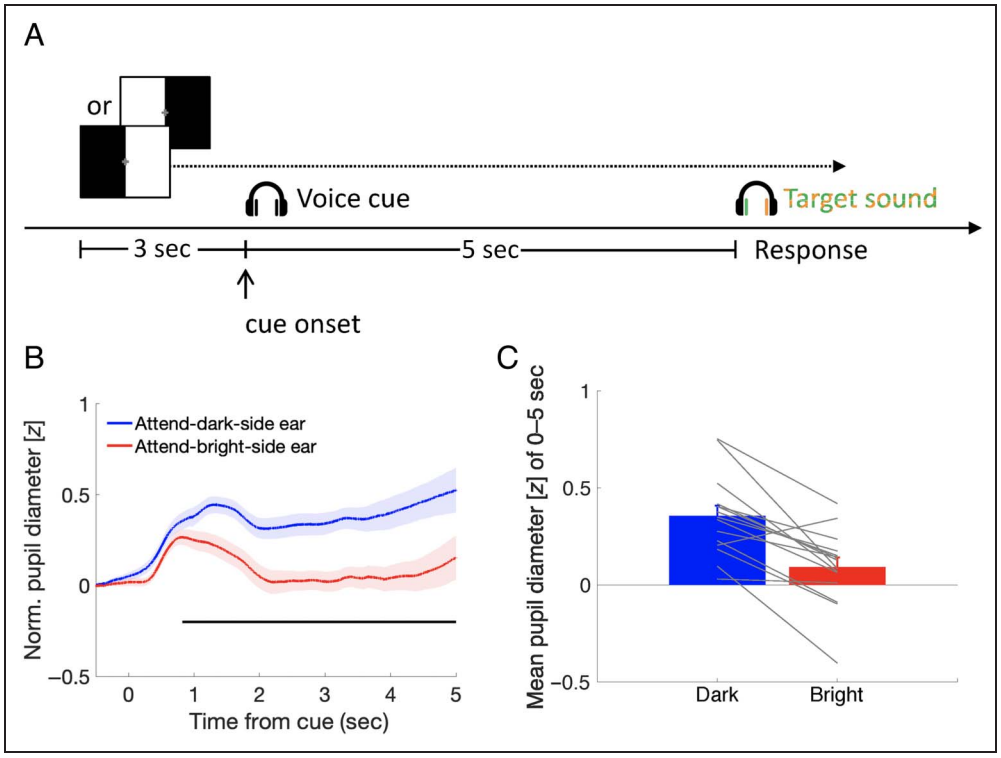

Figure 1. Experimental procedure and pupillary response results in Experiment 1. (A) Schematic procedure. The visual display with two different luminance disparities was randomly presented for each trial. Participants paid attention to their left or right ear, depending on the voice cue, and responded by indicating whether the cued target sound was a "dog barking" or a "phone ringing" as soon and accurately as possible while looking at the central fixation cross for pupillary response recording. (B) Pupillary response as a function of time from the voice cue onset, parameterized with the cue–luminance associations (attend dark or bright side). The shaded area represents standard errors across participants. The horizontal black line indicates a significant difference between the attend-dark and attend-bright conditions (nonparametric cluster-based permutation tests, $p < .05$) during the period of 821–5000 msec. (C) Mean pupil diameter as a function of the cue–luminance associations. Error bars represent standard errors among participants. Gray lines represent data of individual participants. Norm. = Normalized.

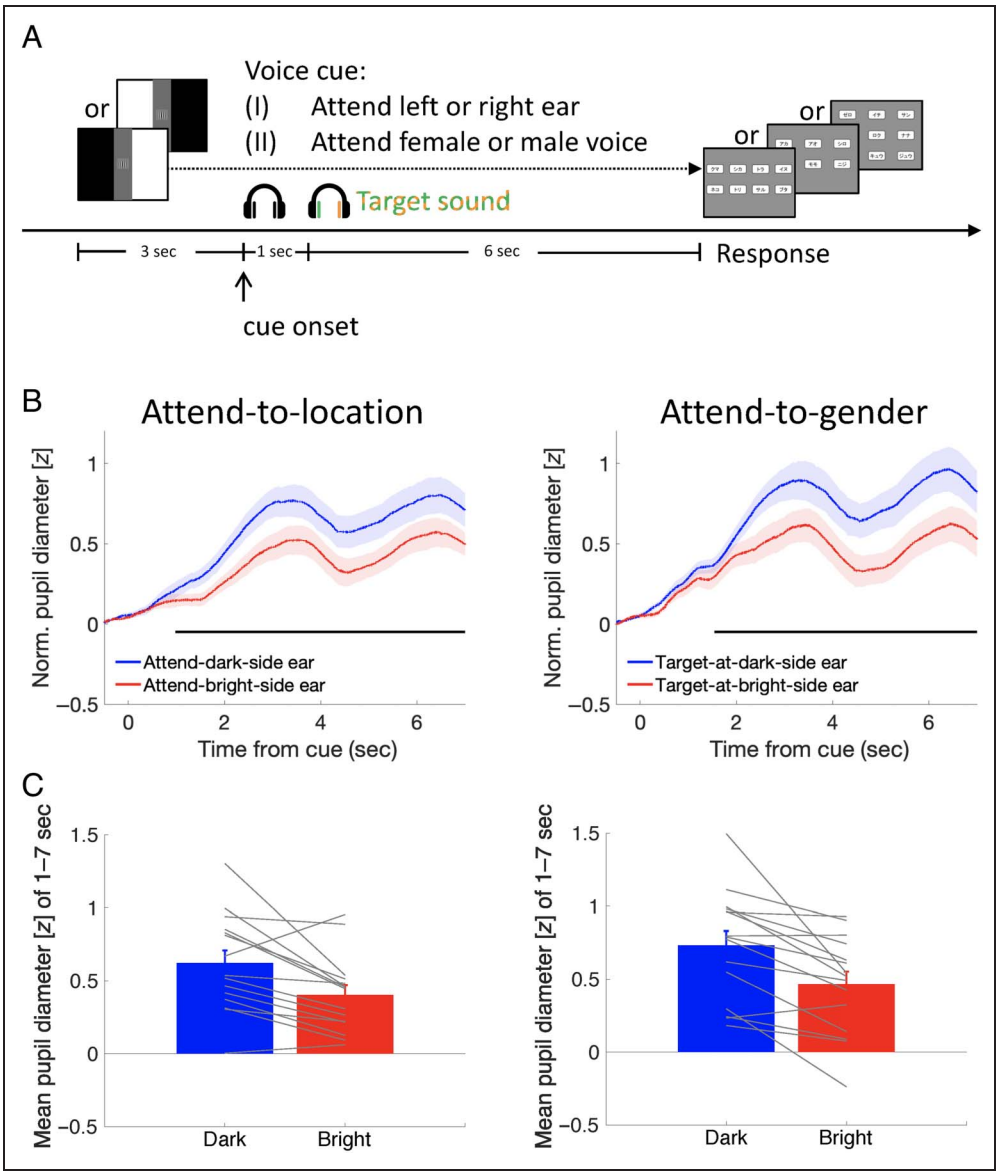

Figure 2. Experimental procedure and pupillary response results in Experiment 2. (A) Schematic of procedure. The visual display with two different luminance disparities was randomly presented for each trial. The center vertical gray band was presented as a "buffer" zone. Participants paid attention to the target sentence, memorized its content, and recalled it after the sound presentation. The target sentence was defined by a spatial cue (I) or a nonspatial cue (II). (B) Pupillary response as a function of time from the voice cue onset in the spatial (i.e., attend-to-location) and nonspatial (i.e., attend-to-gender) cue conditions, parameterized with the cue–luminance associations (attend dark or bright side; left) or the target–luminance associations (target located on the dark or bright side; right), respectively. The shaded area represents standard errors across participants. The horizontal black lines indicate a significant difference between the attend-dark and attend-bright conditions (nonparametric cluster-based permutation tests, $p < .05$) during the period of 978–7000 msec in the left panel and the difference between the target-at-dark and target-at-bright conditions during that of 1542–7000 msec in the right panel. (C) Mean pupil diameter as a function of the cue–luminance associations (left) or target–luminance associations (right). Error bars represent standard errors among participants. Gray lines represent data of individual participants. Norm. = Normalized.

conditions in separate sessions, with the order counterbalanced across participants. Each attention condition consisted of two blocks with a short break in between.

In Experiment 3, the design and procedure were the same as in Experiment 2 except for the following. First, only the gender voice cue (male or female) was used. Second, the target sentences were presented dichotically as well as diotically, where the two target sentences were mixed into one signal and presented to the two ears. The two types of target sound presentation (dichotic, diotic) were presented in randomized order within a block. There were 32 trials for each target sound presentation type, resulting in 64 trials in each block. Participants performed two blocks with a break in between. Experiment 4 followed the same design and procedure of the attend-to-gender condition in Experiment 2, except that all the auditory stimuli were presented via speakers instead of headphones. Participants only performed one block.

Behavioral Data Analyses

In all experiments, trials with incorrect answers were excluded from further analyses (1.7%, 3.4%, 4.4%, and 1.6% of trials excluded in Experiments 1, 2, 3, and 4, respectively). Further exclusion criterion was set for Experiment 1: Trials with an RT exceeding two times the standard deviation from the mean within each session were excluded (4.3% of trials). No RT criterion was set for Experiments 2–4 because participants performed the task with no time pressure.

Eye Metrics Data Analyses

Because eye movements and pupillary responses are consensual, only the data from the left eye were used. During the silent or listening period, blinks accounted for 12.3% of data points in Experiment 1, 12.2% in Experiment 2, 12.8% in Experiment 3, and 12.3% in Experiment 4. The missing pupil-diameter data during blinks were interpolated by using shape-preserving piecewise cubic interpolation. To compare the pupillary response results across participants and conditions, pupil diameter data were normalized by *z*-transform using all the data recorded in each block and then baseline-corrected trial-by-trial by subtracting the mean of the data during the 1-sec period before the cue onset.

Statistical Analysis

Frequentist and Bayesian Statistics

Both frequentist and Bayesian statistics were performed on average data, including RTs (Experiment 1), accuracies, and mean pupil diameters during the silent period (Experiment 1) or the listening period (Experiments 2–4). The Bayesian statistical analyses were computed by using the open-source program JASP (Jeffreys's Amazing Statistic Program; jasp-stats.org), referenced by Keysers, Gazzola, and Wagenmakers (2020). Detailed parameters and tests for each experiment are described below.

Behavioral Performance

In Experiment 1, mean RTs and accuracies were subjected to repeated-measure ANOVAs with Display Disparity (black on left or right), Cue Direction (left, right), and Target Type (dog barking or phone ringing) as within-subject factors. In Experiments 2 and 3, mean accuracies were subjected to paired two-sample *t* tests on the comparison between attend-to-location and attend-to-gender conditions (Experiment 2) and on the comparison between dichotic and diotic conditions (Experiment 3). For between-experiment comparison, mean accuracies in the attend-to-gender condition in Experiment 2 and mean accuracies of the dichotic trials in Experiment 3 were subjected to independent samples *t* test.

Attention-Based PLR Bias

Mean pupil diameters during the silent (i.e., 0–5 sec time-locked to the cue onset in Experiment 1) or listening (i.e., 1–7 sec time-locked to the cue onset in Experiments 2–4) period were subjected to paired two-sample *t* tests with the cue–luminance associations (attend the dark or bright side) for the space cue (Experiments 1 and 2) and with the target–luminance associations (target located at the dark or bright side) for the gender cue (Experiments 2–4) as the independent variables. In Experiment 3, mean pupil diameter data were further subjected to a repeated-measure ANOVA with the Target–Luminance Associations (dichotic trials with the target located on the dark or bright side, or diotic trials with the target interlaced at the center) as a within-subject factor. Between-experiment comparison was conducted as follows: Mean pupil diameters in Experiment 2 (only in the attend-to-gender condition) and Experiment 3 (only the dichotic presentation trials) were subjected to a mixed-design ANOVA with the Target–Luminance Associations (target located at the dark or bright side) as a within-subject factor and Experiment as a between-subject factor.

Eye Metrics Data across Time

Pupil diameter change and luminance of the gazed position were subjected to nonparametric cluster-based permutation tests (Maris & Oostenveld, 2007) to examine the difference between the conditions across time. Luminance contrast of the gazed position was smoothed by moving averaging with a 100-msec time window for each participant. This was done to reduce the noise across time. The cluster-based analyses were computed by using the Fieldtrip MATLAB toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) with 5000 iterations and both the cluster-defining height threshold and FWE-corrected cluster size threshold below an α level of .05.

Linear Mixed-Effects Models

Linear mixed-effects (LME) analyses were performed at each 1-msec sampling point with luminance of the gazed position (dark, bright in Experiment 1; dark, bright, gray in Experiments 2–4) and attended luminance (cue–luminance or target–luminance association as dark, bright) as fixed effects, participant as the random effect, and pupil diameter as the dependent variable. The LME analyses were computed by using the *lem4* package in R (Bates, Mächler, Bolker, & Walker, 2015). Following a criterion similar to Mathôt et al. (2014), the significant clusters were defined as a t value > 2 with at least 500 consecutive samples. Unlike Mathôt et al. (2014), who chose 200 consecutive samples, we set the criterion stricter with larger consecutive samples because the gazed luminance was noisy across time.

PLR Divergence Latency

The latency was defined as the earliest end of the period with the cluster-based significant difference between the attend-to- (or target-at-) dark side and attend-to- (or target-at-) bright side conditions. Following the jackknife resampling technique, the latency of the divergence difference and its variance were estimated. Pupil size difference between the two attended luminance conditions was subjected to nonparametric cluster-based permutation tests by omitting the data from one participant. To guarantee that the effective clusters could be identified for all

participants, the threshold was set to be the smallest α level or the second smallest α level if the difference between them was smaller than .05. The procedure was repeated until all participants were omitted once. The estimated latencies were subjected to both frequentist and Bayesian statistics. Within-experiment comparison was conducted in Experiment 2, in which the estimated latencies were subjected to a paired two-sample t test with the attention condition (attend-to-location, attend-to-gender) as the independent variable. Between-experiment comparisons were conducted for the attend-to-location and attend-to-gender conditions separately. For the attend-to-location condition, the estimated latencies in Experiment 1 and those of the attend-to-location condition in Experiment 2 were subjected to an independent sample t test with the experiment as the grouping variable. For the attend-to-gender condition, the estimated latencies in Experiments 3 and 4 and those of the attend-to-gender condition in Experiment 2 were subjected to ANOVAs with the Experiment as the fixed factor.

RESULTS

Experiment 1: PLR Reflects the Direction of Endogenous Auditory Spatial Attention

We started by replicating the paradigm of Mathôt et al. (2013), but with a modification of the main task to discriminate auditory objects (Experiment 1). Participants listened to an auditory voice cue saying "left" or "right" to shift their attention to the left or right ear, respectively. After 5 sec, two environmental sounds, a dog barking and a phone ringing, were presented for 500 msec dichotically through headphones. They responded by pressing a designated key to report whether the sound presented in the cued ear was "dog barking" or "phone ringing" as soon and accurately as possible (see Figure 1A for procedure).

Behavioral results showed that RTs and accuracies were equally fast and accurate among all the conditions (see Table 1). For RTs, the three-way interaction of Visual Display disparity (darkness at the left or right visual hemifield), Cue Direction (left or right), and Target Sound (dog barking or phone ringing) showed significance, $F(1, 14) = 4.87$, $p = .045$, whereas the Bayes factor (BF) with the model including only the three-way interaction was very small ($BF_{incl} = 0.70$). None of the other effects or interactions were significant (Fs < 3.6, ps $> .08$, $0.1 < BFs_{incl} < 0.6$). For accuracies, none of any main effects or interactions were significant (Fs < 1.8, ps $> .2$, $0.2 < BFs_{incl} < 0.7$).

Most importantly, mean pupil size was larger when the target sound was presented in the cued direction where the visual hemifield was dark than when it was bright, $t(14) = 5.32$, $p < .001$, $BF_{+0} = 530.19$, median $\delta = 1.24$, 95% CI [0.538, 1.976] (Figure 1C). This difference was significant 821 msec after the voice cue onset ($p < .05$, nonparametric cluster-based permutation test; see Figure 1B).

Experiment 2: PLR Reflects Continuous Covert Attention to Auditory Object in Space

In Experiment 1, the spatial attention effect reflected in the PLR was observed during the period in which no auditory stimulus was presented. It could be argued that, during this period, the participant's spatial attention was not directed to an auditory object but to the visual display, and thus the present result is a mere replication of the experiment on visual attention (Mathôt et al., 2013). We examined this possibility in Experiment 2, in which the target sound was changed to a continuous speech sentence presented for 6 sec, and we investigated whether and how the PLR effect was observed during the target sound presentation period. The stimuli for each trial were two sentences randomly sampled from a Japanese version of the CRM corpus (Bolia et al., 2000). Versions of the CRM corpus have been widely used to study speech intelligibility in competing speech or noise, and each sentence in the corpus consists of three keywords, namely, an animal name, color, and number in the Japanese version. In our procedure (see Figure 2A), two sentences by two talkers, one male and one female, were presented dichotically through headphones. In the attend-to-location condition, the voice cue saying "left" or "right" (as in Experiment 1) was presented 1 sec before the target sentence. Participants paid attention to the cued direction (or ear), memorized the keywords in the target sentence presented at the cued direction, and later recalled them. In the attend-to-gender condition, the voice cue said "male" or "female," and participants recalled the content of the sentence expressed by the

Table 1. Mean RTs (msec) and Accuracies (in Parentheses) Under Each Condition in Experiment 1

	Cue-Left		*Cue-Right*	
	Phone	*Dog*	*Phone*	*Dog*
Black-left	1128 (99%)	1149 (97%)	1213 (98%)	1154 (99%)
Black-right	1151 (98%)	1188 (98%)	1144 (99%)	1178 (99%)

cued-gender talker. Participants performed these two conditions in separate blocks, in a counterbalanced order across participants.

Behavioral performance was equally good in these two conditions (mean accuracy = 96.1% and 97.1% for the attend-to-location and attend-to-gender conditions, respectively, $t(14) = 1.81$, $p = .092$, $BF_{10} = 0.97$, median $\delta = -0.396$, 95% CI [−0.916, 0.092]. Critically, in both conditions, mean pupil size was larger when the target sentence appeared in the dark visual hemifield than when it appeared at the bright one: $t(14) = 3.43$, $p = .004$, $BF_{+0} = 23.28$, median $\delta = 0.775$, 95% CI [0.217, 1.385], and $t(14) = 3.97$, $p = .001$, $BF_{+0} = 58.34$, median $\delta = 0.907$, 95% CI [0.306, 1.553], in the attend-to-location and attend-to-gender conditions, respectively (Figure 2C). The difference reached significance 978 msec after the spatial cue (left or right) and 1542 msec after the nonspatial cue (male or female; Figure 2B).

Experiment 3: Does PLR-Reflecting Object-Based Auditory Attention Require Consistent Association with Target Space?

An interesting finding of the above experiment (Experiment 2) was that PLR reflected attention even to a nonspatially guided auditory target. This suggests compulsory involvement of space information in object-based auditory attention. It should be noted, however, that the target sounds were always presented in separate locations (or ears) in that experiment. Thus, it is arguable that the participants might have strategically used the spatial representation of the target sound predominantly even when the target sound was defined by a nonspatial cue and that the observed PLR effectively reflected an intention of attention to space. Experiment 3 was designed to discourage the participant from taking this strategy. We mixed trials with dichotic stimulus presentation (as in Experiment 2) and those with diotic presentation in a random order within a block. In the diotic presentation trial, the two talkers' sentences were mixed into one signal, which was presented to the two ears. In this case, the two voices are not perceptually lateralized, so shifting spatial attention toward the left or right direction would not help in the task. We expected that the modified procedure would encourage the participants to consistently pay more attention to the nonspatial acoustic characteristics to perform the task.

Behavioral results showed that the accuracies of dichotic trials in Experiment 3 were as good as in the attend-to-gender condition in Experiment 2 (96.6% vs. 97.1%, $t(40) = 0.28$, $p = .784$, $BF_{10} = 0.32$, median $\delta = 0.067$, 95% CI [−0.486, 0.633]), and they were better than those of the diotic trials in Experiment 3 (96.6% vs. 93.5%, $t(26) = 6.16$, $p < .001$, $BF_{10} = 11321.85$, median $\delta = 1.115$, 95% CI [0.625, 1.621]).

Critically, mean pupil sizes were different among the three conditions we tested (i.e., dichotic trials with target on the dark side, dichotic trials with target on the bright side, and diotic trials), $F(1.54, 40.15)_{\text{with Greenhouse–Geisser correction}} = 3.74$, $p = .043$, $BF_M = 1.73$. Post hoc multiple comparisons indicated that mean pupil size was larger when the target sound was located in the dark hemifield than when it was located in the bright one ($p_{\text{Bonferroni}} = .026$, $BF_{+0} = 3.11$, median $\delta = 0.389$, 95% CI [0.062, 0.772]), but no significant difference was found for the other two pairs ($ps_{\text{Bonferroni}} > .4$, $BFs_{+0} < 0.94$; Figure 3B). The difference among these conditions was found during the periods of 2036–3403 and 4408–7000 msec (Figure 3A). The segment of two significant periods corresponds to how the target sentences were presented, namely, twice with a short break in between. The result suggests that the allocation of spatial attention to the auditory object matches the temporal dynamic of the auditory object's presentation.

We compared the PLRs in comparable conditions between Experiments 2 and 3, in which the gender-cue-guided target and the two sentences were presented dichotically. The purpose of this comparison was to examine whether the PLR-reflecting object-based attention was modulated by the consistency of the availability of the sound space information. Mean pupil sizes in the dichotic trials in Experiment 3 and those in the attend-to-gender condition in Experiment 2 were subjected to a mixed-design ANOVA with the Target Sound Location (located in the dark or bright hemifield) as a within-subject factor and Experiment as a between-subject factor. Consistent with the previous analyses conducted for individual experiments, mean pupil size was overall larger when the target sound was located in the dark than bright hemifield, $F(1, 40) = 23.22$, $p < .001$, $BF_{\text{incl}} = 112.77$. Mean pupil size did not differ between the two experiments, $F(1, 40) = 0.04$, $p = .849$, $BF_{\text{incl}} = 0.48$. Importantly, the PLR effect was stronger in Experiment 2 than Experiment 3, indexed by the interaction between the target sound location and experiment, $F(1, 40) = 5.75$, $p = .021$, $BF_{\text{incl}} = 2.27$.

The overall results indicate that PLRs reflected spatial auditory attention even when a nonspatial cue guided attention. However, the PLR bias became weaker when the spatial representation of the sound source did not always help for the task at hand (as in Experiment 3).

Experiment 4: Attention-Related PLRs to Sounds from Loudspeakers in Space

It can be argued that the attention-related PLRs reflected the attention directed to the ear, not the space, per se, in the previous three experiments, in which all the sounds were presented via headphones. To test for this possibility, in Experiment 4, we replicated the attend-to-gender condition in Experiment 2 but presented stimuli via loudspeakers, which were located on either side of the visual display. Participants performed the task with the same accuracy as in Experiment 2 (98.4% vs. 97.1%, $t(29) =$

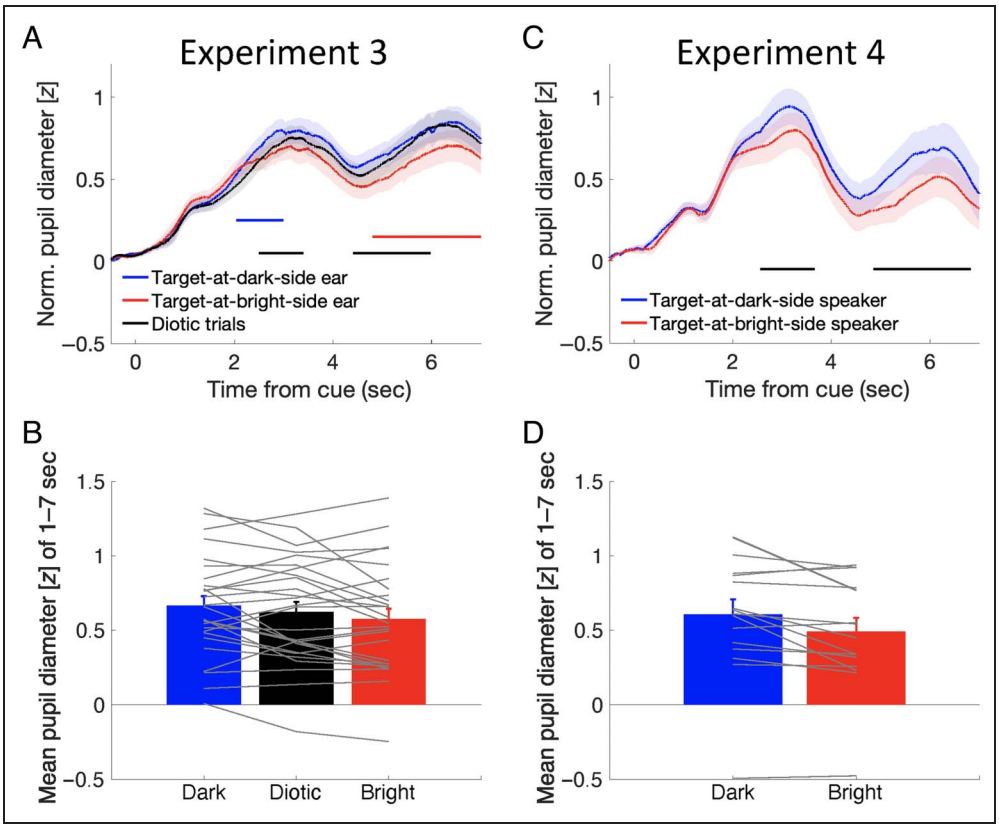

Figure 3. Pupillary response results in Experiments 3 and 4. (A) Pupillary response as a function of time from the voice cue onset, parameterized with the target–luminance associations (dichotic trials with the target located on the dark or bright side, or diotic trials). The shaded area represents standard errors across the participants. The horizontal black line indicates a significant difference between the target-at-dark and target-at-bright conditions during periods of 2493–3403 and 4408–5986 msec. The horizontal blue line indicates a significant difference between the target-at-dark and diotic conditions during the period of 2036–2997 msec. The horizontal red line indicates a significant difference between the target-at-bright and diotic conditions during that of 4803–7000 msec. (B) Mean pupil diameter as a function of the target–luminance associations. Error bars represent standard errors among participants. Gray lines represent data of individual participants. (C) Pupillary response as a function of time from the voice cue onset, parameterized with the target–luminance associations (target located at the dark or bright side). The shaded area represents standard errors across participants. The horizontal black lines indicate a significant difference between the target-at-dark and target-at-bright conditions during the periods of 2559–3669 and 4856–6831 msec. (D) Mean pupil diameter as a function of the target–luminance associations. Error bars represent standard errors among participants. Gray lines represent data of individual participants. Norm. = Normalized.

1.27, $p = .216$, $BF_{10} = 0.62$, median $\delta = -0.340$, 95% CI [−1.016, 0.270]). Critically, the pupillary response results showed that mean pupil size was larger when the target sound came from the loudspeaker located next to the dark visual hemifield than from the one next to the bright one, $t(15) = 3.04$, $p = .008$, $BF_{+0} = 12.62$, median $\delta = 0.666$, 95% CI [0.161, 1.227] (Figure 3D). The difference reached significance 2559–3669 and 4856–6831 msec after the cue presentation (Figure 3C). The segment of the two periods showed a pattern similar to the result in Experiment 3.

Gaze Bias and PLR

It can be argued that the attention-related PLR is because of the change in light inputs after uncontrolled eye movements such as ocular drifts. To address this issue, we first identified the luminance of the local gazed position by matching the gaze position at each 1-msec sampling point and the luminance of the viewed display for each trial. We found that although the global luminance was controlled to be constant, there were significant differences in the local luminance of the gazed position between attending-dark and attending-bright conditions in all experiments. The difference reached significance at 726 msec after the cue presentation in Experiment 1; 494 and 1470 msec in the attend-to-location and attend-to-gender conditions, respectively, in Experiment 2; 1564 msec in Experiment 3; and 1595 msec in Experiment 4 (see Figure 4). One unexpected finding is that the gaze was focused more when a luminance boundary was presented near the fixation as in Experiment 1 (93% of the data points within 2°) than it was in Experiments 2–4 (percentage of data points within the vertical gray area: 67% for the attend-to-location condition and 70% for the attend-to-gender condition in Experiment 2, 66% in Experiment 3, and 74% in Experiment 4). The difference could be also because of the nature of the task. In any case, although the initial motivation for inserting the gray vertical area (2.3 visual degrees in width) in Experiments 2–4 was to make a "buffer zone" between the hemifields to reduce drastic luminance variations in foveal inputs accompanied by small gaze drifts, the gaze analysis suggests that gaze drift was beyond the gray buffer zone and thus still accompanied local luminance variation.

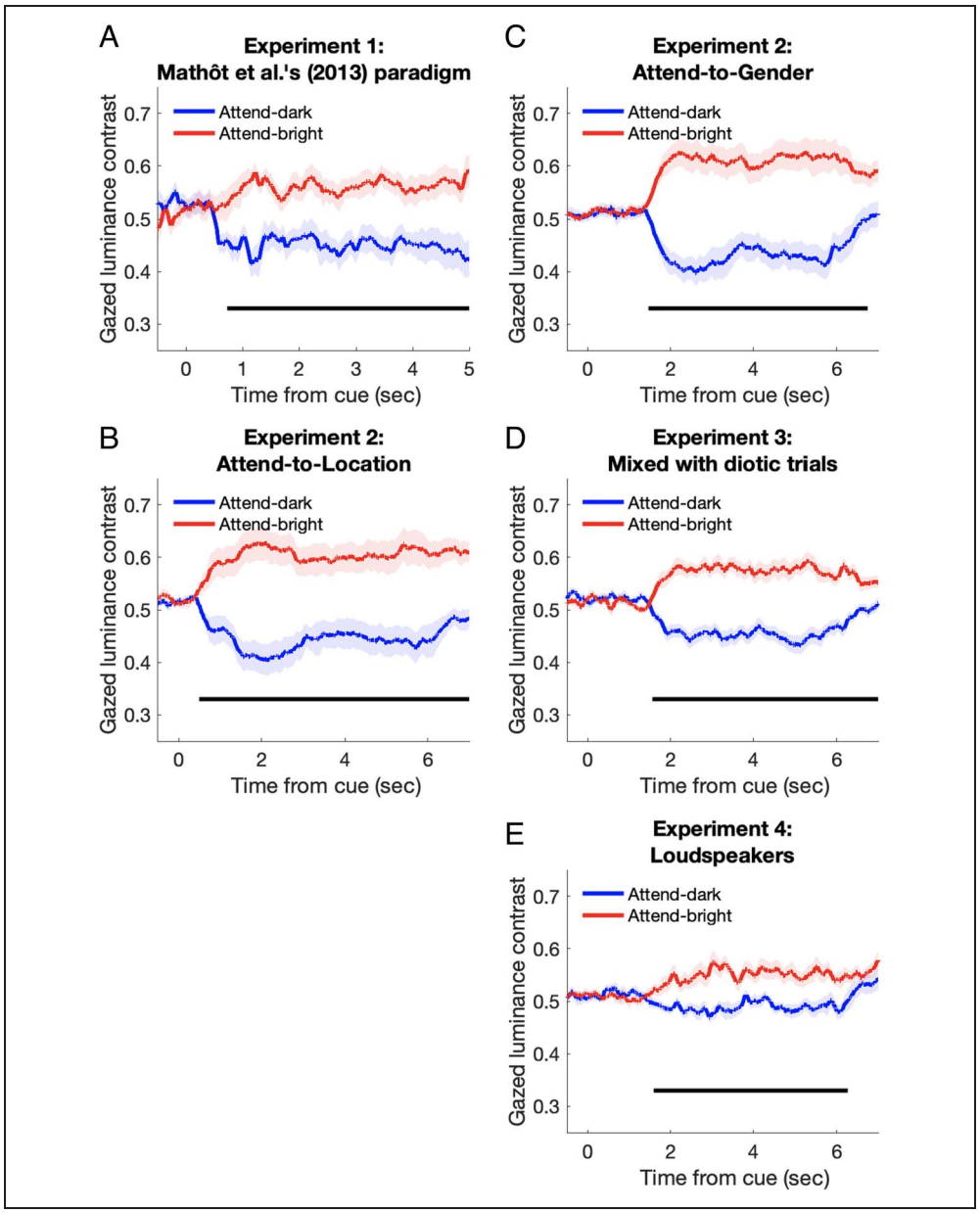

Figure 4. Gazed luminance contrast as a function of time from the voice cue onset, parameterized with the attended luminance condition (dark or bright) in all experiments. The horizontal black lines indicate a significant difference (nonparametric cluster-based permutation tests, $p < .05$) between the two attended luminance conditions during the periods of 726–5000 msec in (A), 494–7000 msec in (B), 1470–6747 msec in (C), 1564–7000 msec in (D), and 1595–6268 msec in (E).

We therefore performed LME analyses to examine whether gazed local luminance or attended luminance could better explain the changes in pupil size. LME analyses were conducted at each 1-msec sampling point with gazed luminance and attended luminance as fixed effects, participant as the random effect, and pupil size as the dependent measure. Results showed that, except for Experiment 3, pupil size was better predicted by attended luminance than gazed luminance. In Experiment 1, only attended luminance, not gazed luminance, significantly predicted pupil size. The t value reached significance 800 msec after the cue onset. In Experiments 2 and 4, attended luminance, compared with gazed luminance, predicted pupil size with earlier timings: 971 versus 1690 msec in the attend-to-location condition, 1861 versus 3848 msec in the attend-to-gender condition in Experiment 2, and 2579 versus 6755 msec in Experiment 4. The significant periods were also longer and more stable for attended luminance than for gazed luminance (see Figure 5). By contrast, in Experiment 3, although attended luminance showed earlier prediction timing than gazed luminance (1977 vs. 2364 msec), the effect did not last long. The gazed luminance instead showed more stable and larger t values than attended luminance. The results suggest that when the spatial representation of the sound source was not always clear and lateralized, as we mixed dichotic trials and diotic trials here, the PLR was better explained by the gazed local luminance than the attended luminance. The uncertainty of the spatial property may evoke unexpected gaze drift, obscuring the attention-evoked effect.

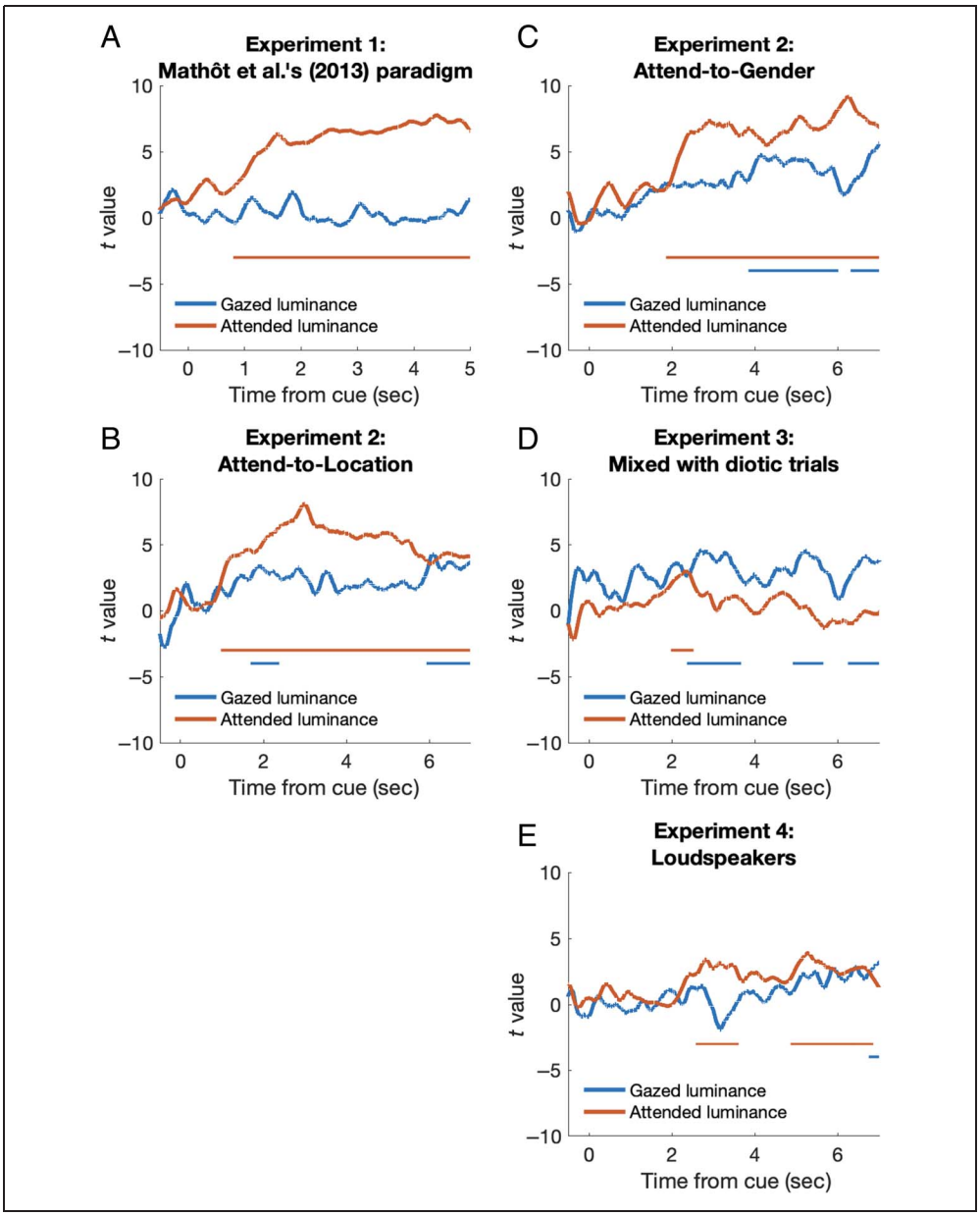

Figure 5. Result of the LME analysis in all experiments. *t* Value as a function of time from the voice cue onset, parameterized with regression variables of gazed luminance (blue lines) and attended luminance (orange lines) predicting pupil size. The plot was smoothed by moving averaging with a 200-msec time window. The horizontal color lines indicate the significant period ($t > 2$) with at least 500 consecutive samples.

Tracking the Time Course of Attentional Shift by PLR Divergence

On the assumption that the instantaneous change in pupil size reflects the state of spatial attention at a time point (with some delay), we may be able to infer the participants' internal process for directing spatial attention. That is, the divergence of the PLR between the attend-to-dark and attend-to-bright conditions can be regarded as reflecting the existence or degree of attention-direction bias. Figure 6 represents the time course of the difference (in z score) between the pupil size for the attend-to- (or target-at-) dark and bright side conditions, summarizing the results of Experiments 1–4. The horizontal color bars at the bottom indicate the time ranges in which the difference (or divergence) was significantly above zero. We adopted the jackknife resampling technique to estimate the variance of the onset latency for statistical analyses (see Methods for details). Results showed that, in Experiment 2, the onset latency was shorter in the attend-to-location condition than in the attend-to-gender condition, $t(14) = 6.38$, $p < .001$, $BF_{10} = 1358.88$, median $\delta = 1.496$, 95% CI [0.724, 2.320]. For the attend-to-location condition, the onset latency was shorter in Experiment 1 than in Experiment 2, $t(28) = 2.74$, $p = .011$, $BF_{10} = 4.89$, median $\delta = 0.814$, 95% CI [0.103, 1.592]. For the attend-to-gender condition, the onset latency was different among the three experiments (i.e., Experiments 2, 3, and 4; $F(2, 55) = 179.00$, $p < .001$, $BF_M = 2.463e{+}21$). Post hoc analysis indicated that the latency was shorter

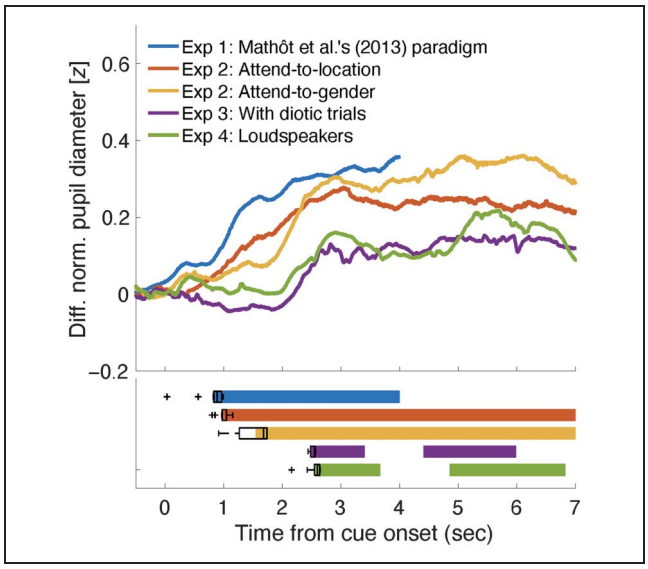

Figure 6. Differences in pupillary responses between the attend-to- (or target-at-) dark and attend-to- (or target-at-) bright conditions, as a function of the time from the cue onset. The horizontal color bars indicate the significance (nonparametric cluster-based permutation tests, $p < .05$) as the pupillary response divergence deviated from zero. The boxplots show the median values, the interquartile range (IQR), IQR × 1.5, and outliers of the divergence latency estimated by the jackknife resampling technique. Exp = Experiment.

in Experiment 2 than in Experiment 3 ($p_{Tukey} < .001$, $BF_{10} = 3.600e{+}15$) or in Experiment 4 ($p_{Tukey} < .001$, $BF_{10} = 3.981e{+}9$).

DISCUSSION

PLR Reflects Spatial Attentional Shift to Auditory Object

In four experiments, we demonstrated that the PLR reflects the focus of covert auditory attention not only when attention is directed to a particular space by an endogenous spatial cue (Experiments 1 and 2) but also when it is allocated to a particular auditory object via nonspatial characteristics (e.g., gender; Experiments 2, 3, and 4). The finding was replicated regardless of whether the sounds were presented via headphones (Experiments 1, 2, and 3) or loudspeakers (Experiment 4).

The time course of the PLR divergence (i.e., the difference between attend-to-dark and attend-to-bright conditions; Figure 6) can be assumed to reflect the timing of the auditory spatial attention shift in participant's planning process of goal-directed action. Given this assumption, we can explain the observed differences in PLR-divergence latencies among the conditions by the following scenarios. When the spatial cue was provided explicitly (i.e., Experiment 1 and the attend-to-location condition in Experiment 2), participants started shifting their attention immediately after the cue presentation. The somewhat longer latency in Experiment 2 than in Experiment 1 may reflect the complexity or difficulty of the task, or the onset of a target sentence, which delays attentional shift. When the cue was nonspatial (i.e., the attend-to-gender condition in Experiments 2, 3, and 4), participants needed extra time to consciously or unconsciously identify the location of the target talker before shifting their spatial attention to the auditory object accordingly. Within a comparable attend-to-gender condition, the latency was longer when the sounds were presented via loudspeakers in space (Experiment 4) than when presented directly to the ears through headphones (Experiment 2). It should be noted that the size of the PLR difference was also smaller in Experiment 4 (loudspeakers) than in Experiment 2 (headphones). The slower and smaller divergence in Experiment 4 may be because of a smaller difference between the two sound sources in the internal representation. With the loudspeaker presentation, the azimuthal angles of the sources were smaller, and the sound from a loudspeaker reached both ears.

The observed PLR effect in Experiment 3 (dichotic and diotic sound presentations mixed in an experimental block in random order) was generally weaker (compared with Experiment 2) and better explained by the local gazed luminance caused by unstable ocular drifting than by attended luminance (Figure 5). This could be because of the complexity of the stimulus presentation procedure, in which the auditory objects were sometimes presented in overlapping space. Participants may have, in general, decreased their incentives for using space representation of the auditory object. Alternatively, they may have changed their strategy trial-by-trial, depending on the availability of a clear space presentation of the auditory object (i.e., left or right vs. center/unknown). Such a trial-by-trial switching of strategy is implied by the behavioral performance in that the accuracy was higher when the sounds were presented in separate spaces (i.e., dichotic trials) than mixed into one signal (i.e., diotic trials). In any case, the Experiment 3 results suggest that, despite the possibility of a coactivation of a spatial-attention-related mechanism by object-based attentional orientation, the spatial attention may more strongly affect fixational eye movements than the PLR per se. These movements can be regarded as ocular drifts or microtremors rather than saccades, because most of the samples of the gaze position fell within 1 visual degree around the fixation point, albeit it is difficult to classify the detailed characteristics with the video-based eye-tracking system used in the current study (Ko, Snodderly, & Poletti, 2016). As a result of the fixational eye movements, together with the nature of the visual display we used here, the local luminance is input to the eyes differently and thus affects pupil sizes. Future study should investigate to use of a wider gray area around the fixation to avoid drastic luminance input differences because of gaze position drifting or to better control the participant's fixation within a designed area by online contingent-gaze position monitoring.

Possible Neural Mechanism of PLR Reflecting Auditory Spatial Attention and Its Implications

The PLR has been considered primarily as a reflex, but it is not entirely accounted for by physical inputs. The neural pathway from the optic nerves to the oculomotor nerves controls the "reflex" response. When the photosensitive ganglion cells in the retina are activated by light, the information is transmitted via the optic nerves to the olivary pretectal nucleus (OPN) and projected to the Edinger–Westphal nucleus (EWN) in the midbrain. The EWN supplies preganglionic parasympathetic fibers to the eye, which exit with the oculomotor nerves and form synapses with neurons in the ciliary ganglion. Postsynaptic fibers of the parasympathetic root leave the ciliary ganglion in short ciliary nerves, which innervate the pupillary sphincter muscle to induce pupil constriction.

In addition to this direct "reflex pathway," several other modulatory inputs exist. Developmentally, the optic nerves are derived from the diencephalon, a division of the forebrain, although they are categorized as cranial nerves involved in sensory and somatic motor functions. Anatomically, the OPN receives inputs from the FEF in pFC (McDougal & Gamlin, 2015; Leichnetz, 1990; Huerta, Krubitzer, & Jon, 1986; Künzle & Akert, 1977). Recent evidence has further demonstrated that subthreshold electrical microstimulation of the FEF modulates the PLR (Ebitz & Moore, 2017). It has been proposed that the PLR is modulated by multiple cortical and subcortical projections, including a direct projection from the FEF to OPN, or indirect projections involving the relayed areas such as occipital visual cortical areas or superior colliculus (SC) to the OPN or EWN (Joshi & Gold, 2020; Binda & Gamlin, 2017).

Recent evidence supports the account that emphasizes the contribution of the SC, especially in relation to orienting responses (Strauch, Wang, et al., 2022; Wang & Munoz, 2018). This SC-centered circuit receives top–down modulations from the FEF, ACC, and lateral intraparietal cortex and projects neural activities to the OPN and EWN, as mentioned above. Microstimulation of the intermediate layers of the SC evokes not only pupil dilation orienting responses (Wang, Boehnke, White, & Munoz, 2012) but also location-specific pupil luminance modulation (Wang & Munoz, 2018). Particularly, Wang and Munoz (2018) used the visual display with constant global luminance, similar to the current study, and demonstrated that through altering activity of the intermediate layers of the SC via electrical stimulation (facilitation) and lidocaine injection (inhibition), pupil size was modulated by local luminance at the next fixated location (i.e., attended location). This serves as the neural basis of top–down modulated pupillary responses to the attended or expected luminance condition. The SC, together with the FEF and lateral intraparietal cortex, constitutes the neural network that takes charge of the control of visual selective attention and eye movements (Maunsell, 2015; Knudsen, 2011). In addition to visual attention, the SC is known to integrate multisensory information and have a cross-modal space map. The inferior colliculus, a major brainstem nucleus in the auditory pathway, transmits information that is essential for forming the space map in the SC (Cohen & Knudsen, 1999). The inferior colliculus also receives descending projections from the auditory cortex and thus could be under the influence of cognitive states, including attention (Huffman & Henson, 1990). Our finding of the PLR's reflecting auditory spatial attention ties closely with the SC-centered account. Assuming that attention-modulated PLR serves for the preparation of the upcoming perception, in particular in the visual domain (Mathôt, 2018; Mathôt & Van der Stigchel, 2015), our finding implies that the spatial representation in auditory space is merged with the pupil-related spatial map in the visual domain.

Conclusions

The current study provides a method to infer auditory spatial attention by examining the luminance condition of the environment and pupillary response. The attention-modulated PLR reflects the dynamics of attentional shift to auditory objects. The overall results imply a unified audiovisual representation of spatial attention. Auditory object-based attention contains a space representation of the attended auditory object, even when the object is oriented without explicit spatial guidance.

Reprint requests should be sent to Hsin-I Liao, NTT Communication Science Laboratories, NTT Corporation, 3-1, Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan, or via e-mail: hsini.liao.pb@hco.ntt.co.jp.

Data Availability Statement

Data and analysis code are available at github.com/hsiniliao/PLR_AuditoryObject.

Author Contributions

H.-I. L. and S. F. developed the study concept, contributed to the study design, and wrote the manuscript. H.-I. L., H. F., and Y.-H. Y. conducted experiments and collected the data. H.-I. L. performed the data analysis. H. F., S. Y. and Y.-H. Y. provided critical comments. All authors approved the final version of the manuscript for submission.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender

identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this article report its proportions of citations by gender category to be as follows: M/M = .707; W/M = .122; M/W = .049; W/W = .122.

REFERENCES

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Best, V., Ozmeral, E. J., Kopčo, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences, U.S.A.*, *105*, 13174–13178. https://doi.org/10.1073/pnas.0803718105, PubMed: 18719099

Best, V., Shinn-Cunningham, B., Ozmeral, E., & Kopco, N. (2010). Exploring the benefit of auditory spatial continuity. *Journal of the Acoustical Society of America*, *127*, EL258–EL264. https://doi.org/10.1121/1.3431093, PubMed: 20550229

Binda, P., & Gamlin, P. D. (2017). Renewed attention on the pupil light reflex. *Trends in Neurosciences*, *40*, 455–457. https://doi.org/10.1016/j.tins.2017.06.007, PubMed: 28693846

Binda, P., Pereverzeva, M., & Murray, S. O. (2013). Attention to bright surfaces enhances the pupillary light reflex. *Journal of Neuroscience*, *33*, 2199–2204. https://doi.org/10.1523/jneurosci.3440-12.2013, PubMed: 23365255

Binda, P., Pereverzeva, M., & Murray, S. (2014). Pupil size reflects the focus of feature-based attention. *Journal of Neurophysiology*, *112*, 3046–3052. https://doi.org/10.1152/jn.00502.2014, PubMed: 25231615

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *Journal of the Acoustical Society of America*, *107*, 1065–1066. https://doi.org/10.1121/1.428288, PubMed: 10687719

Braga, R. M., Fu, R. Z., Seemungal, B. M., Wise, R. J. S., & Leech, R. (2016). Eye movements during auditory attention predict individual differences in dorsal attention network activity. *Frontiers in Human Neuroscience*, *10*, 164. https://doi.org/10.3389/fnhum.2016.00164, PubMed: 27242465

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357, PubMed: 9176952

Bushara, K. O., Weeks, R. A., Ishii, K., Catalan, M.-J., Tian, B., Rauschecker, J. P., et al. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nature Neuroscience*, *2*, 759–766. https://doi.org/10.1038/11239, PubMed: 10412067

Cave, K. R., & Bichot, N. P. (1999). Visuospatial attention: Beyond a spotlight model. *Psychonomic Bulletin & Review*, *6*, 204–223. https://doi.org/10.3758/BF03212327, PubMed: 12199208

Cohen, Y. E., & Knudsen, E. I. (1999). Maps versus clusters: Different representations of auditory space in the midbrain and forebrain. *Trends in Neurosciences*, *22*, 128–135. https://doi.org/10.1016/S0166-2236(98)01295-8, PubMed: 10199638

Corbetta, M. (1998). Frontoparietal cortical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences, U.S.A.*, *95*, 831–838. https://doi.org/10.1073/pnas.95.3.831, PubMed: 9448248

Downing, C. J. (1988). Expectancy and visual–spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 188–202. https://doi.org/10.1037/0096-1523.14.2.188, PubMed: 2967876

Ebitz, R. B., & Moore, T. (2017). Selective modulation of the pupil light reflex by microstimulation of prefrontal cortex. *Journal of Neuroscience*, *37*, 5008–5018. https://doi.org/10.1523/jneurosci.2433-16.2017, PubMed: 28432136

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*, 225–240. https://doi.org/10.3758/BF03211502, PubMed: 3786090

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention—Focusing the searchlight on sound. *Current Opinion in Neurobiology*, *17*, 437–455. https://doi.org/10.1016/j.conb.2007.07.011, PubMed: 17714933

Haab, O. (1886). Vortrag gesellschaft der ärzte in zürich am 21. November 1885. *Korrespondenzblatt für Schweizer Ärzte*, *16*, 153.

Hill, K. T., & Miller, L. M. (2009). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex*, *20*, 583–590. https://doi.org/10.1093/cercor/bhp124, PubMed: 19574393

Huerta, M. F., Krubitzer, L. A., & Jon, H. C. K. (1986). Frontal Intracortical eye field as defined by microstimulation in squirrel monkeys, owl monkeys, and macaque monkeys: I. subcortical connections. *Journal of Comparative Neurology*, *253*, 415–439. https://doi.org/10.1002/cne.902530402, PubMed: 3793998

Huffman, R. F., & Henson, O. W. (1990). The descending auditory pathway and acousticomotor systems: Connections with the inferior colliculus. *Brain Research Reviews*, *15*, 295–323. https://doi.org/10.1016/0165-0173(90)90005-9, PubMed: 2289088

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

Keysers, C., Gazzola, V., & Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*, 788–799. https://doi.org/10.1038/s41593-020-0660-4, PubMed: 32601411

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 1–16.

Knudsen, E. I. (2011). Control from below: The role of a midbrain network in spatial attention. *European Journal of Neuroscience*, *33*, 1961–1972. https://doi.org/10.1111/j.1460-9568.2011.07696.x, PubMed: 21645092

Ko, H.-K., Snodderly, D. M., & Poletti, M. (2016). Eye movements between saccades: Measuring ocular drift and tremor. *Vision Research*, *122*, 93–104. https://doi.org/10.1016/j.visres.2016.03.006, PubMed: 27068415

Künzle, H., & Akert, K. (1977). Efferent connections of cortical, area 8 (frontal eye field) in *Macaca fascicularis*. A reinvestigation using the autoradiographic technique. *Journal of Comparative Neurology*, *173*, 147–164. https://doi.org/10.1002/cne.901730108, PubMed: 403205

Laeng, B., & Endestad, T. (2012). Bright illusions reduce the eye's pupil. *Proceedings of the National Academy of Sciences, U.S.A.*, *109*, 2162–2167. https://doi.org/10.1073/pnas.1118298109, PubMed: 22308422

Laeng, B., & Sulutvedt, U. (2014). The eye pupil adjusts to imaginary light. *Psychological Science*, *25*, 188–197. https://doi.org/10.1177/0956797613503556, PubMed: 24285432

Larson, E., & Lee, A. K. C. (2014). Switching auditory attention using spatial and non-spatial features recruits different cortical networks. *Neuroimage*, *84*, 681–687. https://doi.org/10.1016/j.neuroimage.2013.09.061, PubMed: 24096028

Leichnetz, G. R. (1990). Preoccipital cortex receives a differential input from the frontal eye field and projects to the pretectal olivary nucleus and other visuomotor-related structures in the rhesus monkey. *Visual Neuroscience*, *5*, 123–133. https://doi.org/10.1017/S095252380000016X, PubMed: 2177637

Liao, H.-I., Kashino, M., & Shimojo, S. (2021). Attractiveness in the eyes: A possibility of positive loop between transient pupil constriction and facial attraction. *Journal of Cognitive Neuroscience*, *33*, 315–340. https://doi.org/10.1162/jocn_a_01649, PubMed: 33166194

Liao, H.-I., Kidani, S., Yoneya, M., Kashino, M., & Furukawa, S. (2016). Correspondences among pupillary dilation response, subjective salience of sounds, and loudness. *Psychonomic Bulletin & Review*, *23*, 412–425. https://doi.org/10.3758/s13423-015-0898-0, PubMed: 26163191

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024, PubMed: 17517438

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

Mathôt, S., Dalmaijer, E., Grainger, J., & Van der Stigchel, S. (2014). The pupillary light response reflects exogenous attention and inhibition of return. *Journal of Vision*, *14*, 7. https://doi.org/10.1167/14.14.7, PubMed: 25761284

Mathôt, S., van der Linden, L., Grainger, J., & Vitu, F. (2013). The pupillary light response reveals the focus of covert visual attention. *PLoS One*, *8*, e78168. https://doi.org/10.1371/journal.pone.0078168, PubMed: 24205144

Mathôt, S., & Van der Stigchel, S. (2015). New light on the mind's eye: The pupillary light response as active vision. *Current Directions in Psychological Science*, *24*, 374–378. https://doi.org/10.1177/0963721415593725, PubMed: 26494950

Maunsell, J. H. R. (2015). Neuronal mechanisms of visual attention. *Annual Review of Vision Science*, *1*, 373–391. https://doi.org/10.1146/annurev-vision-082114-035431, PubMed: 28532368

McDougal, D. H., & Gamlin, P. D. (2015). Autonomic control of the eye. *Comprehensive Physiology*, *5*, 439–473. https://doi.org/10.1002/cphy.c140014, PubMed: 25589275

Mondor, T. A., & Zatorre, R. J. (1995). Shifting and focusing auditory spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 387–409. https://doi.org/10.1037//0096-1523.21.2.387, PubMed: 7714479

Naber, M., Frässle, S., & Einhäuser, W. (2011). Perceptual rivalry: Reflexes reveal the gradual nature of visual awareness. *PLoS One*, *6*, e20910. https://doi.org/10.1371/journal.pone.0020910, PubMed: 21677786

Naber, M., & Nakayama, K. (2013). Pupil responses to high-level image content. *Journal of Vision*, *13*, 7. https://doi.org/10.1167/13.6.7, PubMed: 23685390

Noyce, A. L., Kwasa, J. A. C., & Shinn-Cunningham, B. G. (2022). Defining attention from an auditory perspective. *WIREs Cognitive Science*, e1610. https://doi.org/10.1002/wcs.1610, PubMed: 35642475

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869, PubMed: 21253357

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. https://doi.org/10.1163/156856897X00366, PubMed: 9176953

Posner, M., Snyder, C., & Davidson, B. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, 160–174. https://doi.org/10.1037/0096-3445.109.2.160, PubMed: 7381367

Quinlan, P. T., & Bailey, P. J. (1995). An examination of attentional control in the auditory modality: Further evidence for auditory orienting. *Perception & Psychophysics*, *57*, 614–628. https://doi.org/10.3758/BF03213267, PubMed: 7644322

Rhodes, G. (1987). Auditory attention and the representation of spatial information. *Perception & Psychophysics*, *42*, 1–14. https://doi.org/10.3758/BF03211508, PubMed: 3658631

Salvaggio, S., Andres, M., Zénon, A., & Masson, N. (2022). Pupil size variations reveal covert shifts of attention induced by numbers. *Psychonomic Bulletin & Review*, *29*, 1844–1853. https://doi.org/10.3758/s13423-022-02094-0, PubMed: 35384595

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*, 182–186. https://doi.org/10.1016/j.tics.2008.02.003, PubMed: 18396091

Smith, D. V., Davis, B., Niu, K., Healy, E. W., Bonilha, L., Fridriksson, J., et al. (2010). Spatial attention evokes similar activation patterns for visual and auditory stimuli. *Journal of Cognitive Neuroscience*, *22*, 347–361. https://doi.org/10.1162/jocn.2009.21241, PubMed: 19400684

Spence, C. J., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 555–574. https://doi.org/10.1037/0096-1523.20.3.555

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1005–1030, https://doi.org/10.1037/0096-1523.22.4.1005, PubMed: 8756965

Strauch, C., Romein, C., Naber, M., Van der Stigchel, S., & Ten Brink, A. F. (2022). The orienting response drives pseudoneglect—Evidence from an objective pupillometric method. *Cortex*, *151*, 259–271. https://doi.org/10.1016/j.cortex.2022.03.006, PubMed: 35462203

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

Suzuki, Y., Minami, T., Laeng, B., & Nakauchi, S. (2019). Colorful glares: Effects of colors on brightness illusions measured with pupillometry. *Acta Psychologica*, *198*, 102882. https://doi.org/10.1016/j.actpsy.2019.102882, PubMed: 31288107

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136. https://doi.org/10.1016/0010-0285(80)90005-5, PubMed: 7351125

Wang, C. A., Boehnke, S. E., White, B. J., & Munoz, D. P. (2012). Microstimulation of the monkey superior colliculus induces pupil dilation without evoking saccades. *Journal of Neuroscience*, *32*, 3629–3636. https://doi.org/10.1523/jneurosci.5512-11.2012, PubMed: 22423086

Wang, C.-A., & Munoz, D. P. (2018). Neural basis of location-specific pupil luminance modulation. *Proceedings of the National Academy of Sciences, U.S.A.*, *115*, 10446–10451. https://doi.org/10.1073/pnas.1809668115, PubMed: 30249636

Zatorre, R. J., Mondor, T. A., & Evans, A. C. (1999). Auditory attention to space and frequency activates similar cerebral systems. *Neuroimage*, *10*, 544–554. https://doi.org/10.1006/nimg.1999.0491, PubMed: 10547331