Human voices escape the auditory attentional blink: Evidence from detections and pupil responses

Merve Akça a,b, Laura Bishop a,b,1, Jonna Kataariina Vuoskoski a,b,c, Bruno Laeng a,c,*,1

a RITMO Center for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Norway
b Department of Musicology, University of Oslo, Norway
c Department of Psychology, University of Oslo, Norway

ARTICLE INFO

Keywords:
Locus coeruleus-noradrenaline
Temporal selective attention
Auditory attention
Pupillometry

ABSTRACT

Attentional selection of a second target in a rapid stream of stimuli embedding two targets tends to be briefly impaired when two targets are presented in close temporal proximity, an effect known as an attentional blink (AB). Two target sounds (T1 and T2) were embedded in a rapid serial auditory presentation of environmental sounds with a short (Lag 3) or long lag (Lag 9). Participants were to first identify T1 (bell or sine tone) and then to detect T2 (present or absent). Individual stimuli had durations of either 30 or 90 ms, and were presented in streams of 20 sounds. The T2 varied in category: human voice, cello, or dog sound. Previous research has introduced pupillometry as a useful marker of the intensity of cognitive processing and attentional allocation in the visual AB paradigm. Results suggest that the interplay of stimulus factors is critical for target detection accuracy and provides support for the hypothesis that the human voice is the least likely to show an auditory AB (in the 90 ms condition). For the other stimuli, accuracy for T2 was significantly worse at Lag 3 than at Lag 9 in the 90 ms condition, suggesting the presence of an auditory AB. When AB occurred (at Lag 3), we observed smaller pupil dilations, time-locked to the onset of T2, compared to Lag 9, reflecting lower attentional processing when ‘blinking’ during target detection. Taken together, these findings support the conclusion that human voices escape the AB and that the pupillary changes are consistent with the so-called T2 attentional deficit. In addition, we found some indication that salient stimuli like human voices could require a less intense allocation of attention, or noradrenergic potentiation, compared to other auditory stimuli.

When two targets are presented in close temporal succession within a rapid serial presentation of stimuli, people tend to report the second target (T2) incorrectly after correctly identifying the first target (T1), an effect known as the attentional blink (AB; Raymond et al., 1992). A fundamental cause of AB may be difficulty in making rapid attention adjustments, or more specifically, engaging attention twice within a very short period of time (Nieuwenstein et al., 2009).

The present study focuses on evidence that the AB to certain social stimuli can be less affected or ‘attenuated’ than for other stimulus categories. The attenuation of AB for faces has been suggested to represent something like a ‘pop-out effect’ (i.e., pre-attentive processing) for face stimuli or evidence for multiple bottlenecks of information processing (e.g., Awh et al., 2004; Landau and Bentin, 2008). Alternatively, familiarity with faces (‘face expertise’; e.g., Gauthier and Nelson, 2001) may place fewer demands on the limited attentional resources (Jackson and Raymond, 2006). Based on the established cognitive and neural similarities between face and human voice processing (Yovel and Belin, 2013; Belin, 2017), we propose that voices, in general, may be less susceptible to auditory AB effects, similarly to the reduction of AB for faces in the visual modality. In addition to being salient and highly familiar stimuli that are particularly relevant for human listeners, data from behavioral (e.g., Agus et al., 2012; Suied et al., 2014) and electrophysiological studies (e.g., Charest et al., 2009; Bruneau et al., 2013) suggest a rapid and effortless processing of human voices.

Indeed, in a previous study, using an auditory AB task, Akça et al. (2020) investigated the likelihood of presence or absence of an auditory AB effect induced by human voices, a common expertise area for most human listeners. Specifically, by manipulating target sound category (T2), this study compared detecting human voices with that for cello and organ tones (control conditions) in expert cellists and of non-musician.
participants. The evidence indicated the absence of an AB effect for human voices as compared with the other conditions, independently of the participants’ musical expertise.

The present study extends the previous study (Akça et al., 2020) by showing that human voices, when presented as T2, within rapid serial auditory presentation streams, are less likely to evoke an auditory AB effect than the sounds of a cello or dog barking. Cello sounds were selected for two main reasons: First, they share perceptual similarity with human voices (Askenfelt, 1991; Levy et al., 2003) but are not biological sounds. Second, previously reported evidence for the lack of an AB effect in response to cello sounds remained inconclusive (Akça et al., 2020). Dog sounds, on the other hand, differ greatly from human voices acoustically, but are also biological and highly familiar to human listeners, similarly to human voices. Contrasting (behavioural and pupillary) responses to these selected T2 categories will help us better understand whether certain shared characteristics could help explain resilience to the AB effect for human voices and elucidate the differences in attentional allocation deployed when selectively attending to them.

Next, we asked whether stimulus duration has an influence on auditory AB effects triggered by human voices (vs. other sound categories). As shown by previous auditory AB studies (e.g., Arnell and Jolicœur, 1999; Shen and Alain, 2010), presentation rate and stimulus duration can both have a dramatic impact on producing a reliable AB. It is possible that in the previous study (Akça et al., 2020), there was no reliable AB because the stimulus duration was too long (i.e., 150 ms) and in turn the presentation rate (i.e., 160 ms/item) was too slow to produce the AB effect. In the present study, we presented the stimuli in two briefer durations (30 ms vs 90 ms), which also resulted in faster presentation rates (40 ms/item vs 100 ms/item). Indeed, it is typical in visual AB studies to use a stimulus duration that is very brief yet still allows for above chance level target recognition during the rapid serial auditory presentation paradigm. In another study (Akça, Vuoskoski, Laeng, and Bishop, submitted), we confirmed that for auditory stimuli presented for only 30 ms, the recognition sensitivity of a single target during RSAP was well above chance level (Mean d’ = 1.77 ± 0.79 SD).

In addition, Nieuwenhuis et al. (2005) have proposed an influential neurocomputational model of the mechanisms underlying AB. Their model ties AB to the temporal dynamics of the locus coeruleus-noradrenaline (LC-NA) system (i.e., the neurobiological mechanism that causes the release of norepinephrine via neurons located in locus coeruleus) and the occurrence of a refractory-like period following a neural response to T1.

Specifically, according to Nieuwenhuis and colleagues’ (Nieuwenhuis et al., 2005) phasic responses of the locus coeruleus facilitate registering concurrent task-relevant or salient stimuli by modulating the activity of the whole cortex at critical points in time (see also, Poe et al., 2020). Following its neuromodulation or potentiation, the LC neurons go into a brief refractory-like period, like all neurons. The authors proposed that this temporary unavailability of the LC activity might be the main neural process that is responsible for the AB effect. However, several studies show evidence that there exist “non-blinders” in typical visual AB paradigms; that is, around 5% of individuals show little or no AB effect (Martens et al., 2006; Martens and Valchev, 2009; Martens and Johnson, 2009). Moreover, some stimuli (e.g., personal names, faces, voices, expertise objects) may be less likely to be affected by such temporal T2 processing deficit. To account for these “exceptions”, Nieuwenhuis and colleagues (Nieuwenhuis et al., 2005) suggest that certain stimuli (e.g., one’s name) require minimal attentional processing to reach detection threshold and in such cases noradrenergic potentiation might not be crucial. A similar reasoning can account for how certain stimuli can “escape” the so-called T2 deficit, as in our topic of interest, the potential absence or at least attenuation of the AB effect for voices.

Most important for the present study, it is possible to use pupillary responses as an indirect index of the LC-NA system’s activity since there is strong evidence for a close association between pupillary response and LC-NA activity (both in humans, monkeys, and rodents; Alnæs et al., 2014; Joshi and Gold, 2020; Joshi and Gold, 2022; Liu et al., 2017). Pupil dilations occur in response to the detection of a single target during rapid serial presentation in a visual AB task (Privitera et al., 2010) and pupillometry can be a useful marker of the amount of attention deployed to target events also in a rapid serial presentation streams (Zylberberg et al., 2012), as typically used in studying the attentional blink. In the field of hearing science, pupillometry is commonly used to study “listening effort” (i.e., allocation of cognitive resources during listening, Zekveld et al., 2018) associated with, for example, speech intelligibility, linguistic processing, and verbal memory load (e.g., Winn et al. (2018); Zekveld et al.). Thus, the pupil dilation response can also be used as an index for attentional capture by sounds that are both task-relevant and task-irrelevant.

Finally, as mentioned, Willems and Martens (2016) point out that there are large individual differences in AB task performance, with some individuals even showing no AB in certain paradigms. Thus, in the present study, we explored two factors related to the participants that may influence the allocation of attention and the AB effect: impulsivity and musicality. Trait impulsivity (i.e., the tendency to act with little forethought, Dickman, 1990) is theoretically based on individual differences in information processing. Differences in information-processing speed and accuracy are also relevant in the context of AB.

Impulsivity has been previously found to be associated with the visual AB (Li et al., 2005; Troche and Rammayer, 2013) but the results are rather mixed regarding which aspect of impulsivity could be driving the results in Li and colleagues’ Li et al. (2005) study. As for musicality, our previous findings indicated a general musical expertise advantage in the AB task when comparing expert cellists versus novices. To better understand the musicality effects, here we explore musicality in a much broader sense by using a measure of musicality that accounts for different aspects of musical activity (e.g., listening habits, emotional engagement) in addition to performance expertise.

0.1. Aim of the study

The present study aims to a) replicate and further clarify the occurrence of an AB ‘attenuation’ effect with the human voice as a target in an auditory temporal selective attention task and b) consider a model that links LC activity to the AB phenomenon by examining pupillary changes in response to stimulus-related factors (i.e., T2 type, duration and lag). Hence, we performed a pupillometry experiment during an auditory version of the attentional blink paradigm, in which participants were asked to report auditory targets (T1 and T2) presented among auditory distractors in a rapid auditory presentation stream at varying stimulus onset asynchronies (SOAs) or lags between T1 and T2. Thus, in addition to behavioural performance in the AB task, by investigating pupillary changes time-locked to the onset of T2, we aim to reveal the fluctuations in attentional deployment during the auditory AB task and for the different T2 types.

0.2. Hypotheses and predictions

For the behavioral effects, we make the following hypotheses and predictions:

**H1**: We hypothesize that the salience or behavioral significance of the second target (T2) influences its temporal attentional filtering. Hence, the auditory AB effect (as measured with T2–T1 accuracy) should be the least likely to occur for T2s that are higher in salience. Based on our previous findings (e.g., Akça et al., 2020), we expect that there will be an absent or at least attenuated AB effect when T2 is a human voice intoning a vowel sound, compared to the cello tone and dog bark. However, if this temporal attentional advantage for human voices was to extend to sounds that share perceptual similarity with human voices, we would expect to see an attenuated AB effect also when T2 is a cello sound. All considered, we expect that the AB effect is mostly likely to occur when T2 is a dog sound. Finding instead that the AB is
more likely to occur for cello sounds than dog sounds could suggest alternatively that biological and familiar sounds rather than acoustical similarity play a crucial role in the AB.

**H2:** Given the data-limited conditions, if AB is more likely to be present under faster than slower rates, we expect lower overall task performance under 30 than 90 ms condition. However, the typical presentation rate used in AB studies (10 Hz) is equivalent to the 90 ms condition (-10 ms ISI) in the present study. If the AB effect is more likely to occur under this rate (as found by some visual AB studies, e.g., Shapiro et al., 2017), then we would not expect necessarily to find an AB effect in the 30 ms condition, since this condition corresponds to 25 Hz rate which is outside the frequency ‘sweet spot’ mentioned earlier.

Regarding the participant-factors of impulsivity and musical sophistication:

**H3:** Consistent with the previous findings where auditory AB effects appear to be more attenuated in musicians than in non-musicians (Sła-wiński et al., 2002; Martens et al., 2015; Akça et al., 2020), and based on the reported association between impulsivity and the visual AB performance (e.g., Troche and Rammsayer, 2013), we expect that participants’ general musical sophistication scores as well as functional impulsivity scores will negatively correlate with the auditory AB magnitude.

For the pupillary response:

**H4:** We hypothesize that pupil dilation would index the cognitive load (effort) associated with the allocation of attention to specific targets. As suggested by the neurocomputational account of AB (Nieuwenhuis et al., 2005), LC neurons are in a refractory period during AB, so that during AB there should be lower or no phasic pupil dilations. Consistently, the pupillometry literature on target detection (e.g., Privitera et al., 2010) and AB (e.g., Wierda et al., 2012) research suggests that while a successful detection results in phasic dilations, the AB reduces this pupil response. Hence, we expect larger pupils when the targets are likely to be detected versus relatively smaller pupils when T2 remains undetected. Note that, according to these predictions, the pupil is not simply another index of the likelihood of errors (as originally remarked by Kahneman, 1973, p. 25) and it is to be expected that within the AB paradigm, there will be a transient or momentary effort to process and analyze the target stimuli also when behaviorally there is no trace of a T2 detection deficit. Indeed, a variety of pupillometry studies have shown that attention is mobilized in response to the changing demands of a task as one engages in it (Kahneman, 1973, p. 26), so that even when considering only accurate responses - the momentary effort fluctuations due to the allocation of attention will be indexed by changes in the size of pupil dilations. For example, as shown in the seminal studies by Kahneman and Beatty (1966) or by Hess and Polt (1964), pupil size scales with the amount of processing or cognitive load (effort) for all accurate responses (i.e., successful digit memory or mathematical solutions, respectively). Importantly, even attending to a single target tone causes a small and brief increase in pupil size, whereas unattended tones result in no response (Beatty, 1982, p. 284), indicating that the simple detection of a specific target poses demands on processing resources. Finally, we expect that participants’ general musical sophistication scores as well as functional impulsivity scores will negatively correlate with the average pupillary response of each participant.

1. Method

1.1. Participants

Fifty-seven participants with normal or corrected-to-normal (with soft contact lenses) eyesight and normal hearing (by self-report) took part in the experiment in exchange for a gift card worth 200 NOK. None of the participants reported a history of neurological disorders. One participant was excluded from the analyses because of technical problems with the eye tracker which resulted in incomplete pupil and behavioral data. The final sample consisted of 56 participants (38 females, 18 males; mean age = 24.66; range = 19 – 41 years). Expecting a small/moderate effect size of 0.4 (i.e., the typical effect size observed in psychology), would require a sample of at least 50 participants with 80% power and the alpha level of 0.05 (Brynsbaer, 2019). All participants signed a written informed consent form before participation. The protocol conformed to the Helsinki Declaration and to the national guidelines for studies involving human participants.

1.2. Auditory stimuli

The stimuli consisted of snippets of sounds from the categories of human voice, cello, dog, bell, sine tone, and environmental sounds. We sampled the cello and bell sounds from the McGill University Master Samples DVD Set (Opolski and Wapnick, 2006), voices from the Berklee College of Music Sampling Archive Vol. 5 (Boulanger, 2007), and dog bark sounds from the Environmental Sound Classification (ESC-50) dataset (Piczak, 2015). The remainder of the stimuli were sampled from freesound.org.

Within the experimental paradigm, sine tones and bell sounds served as T1, while the human voice, cello, dog sounds served as T2. Environmental sounds were used as distractors and included sounds of everyday objects such as a printer, motorcycle, etc. Human voices comprised recordings of the vowel/a/ sung by a female singer. The targets were selected at 3 different pitches (D4–311 Hz, F4–349 Hz, A4–440 Hz). All stimuli were normalized with peak normalization method and truncated to 30 and 90 ms (with 2-ms linear amplitude ramps) from the quasi-stationary portions of the sounds. In another report (Akça, Vuoskoski, Laeng, and Bishop, submitted) which used the same stimuli as the current study, we showed evidence that these durations were sufficient for accurate recognition of the stimuli (e.g., human voice, cello, dog, etc.). Timbre-related spectral features of the stimulus categories are described in Supplementary Material A and illustrated in Supplementary Fig. B3.

1.3. Self-report inventories

Goldsmith Musical Sophistication Index (Gold-MSI). The Gold-MSI (Müllensiefen et al., 2014) was employed to quantify participants’ self-reported musical sophistication, a multifaceted construct that refers to musical skills, expertise, achievements, and related behaviors. The self-report inventory consists of 38 items (α = 0.93), each rated on a seven-point response scale, plus an open-ended question regarding the tendency is optimal. Together, these 23 items formed the DII-short. The responses to each statement in the scale were given on a True/False format with higher scores indicating higher musical sophistication.

Dickman’s Impulsivity Inventory, short version (DII-short). The DII (Dickman, 1990) was used to assess self-reported trait impulsivity. The scale distinguishes dysfunctional (12 items; score range: 0 – 12; α = 0.84) and functional (11 items; score range: 0 – 11; α = 0.74) components of impulsivity, where the former refers to the tendency to act with little forethought (hence, engaging in a rapid but non-optimal, error-prone information processing) and the latter refers to when this tendency is optimal. Together, these 23 items formed the DII-short. The responses to each statement in the scale were given on a True/False format with higher scores indicating higher impulsivity.

1.4. Apparatus and setup

Auditory stimuli were played at a comfortable listening level over Beyerdynamic DT770 Pro circumaural headphones. Pupil diameter was measured using a binocular remote eye tracking device (RED) by SensoMotoric Instruments (SMI, Teltow, Germany) positioned at 70 cm viewing distance. Pupil size was measured at a sampling rate of 60 Hz from both eyes. Chin rest with forehead support was used for head
movement stabilization. We used SMI’s Experiment Center software to present the stimuli and record pupil data, and BeGaze software for event analysis. We created a fixation image (i.e., fixation cross) for baseline measure and an image of a small circle for the trials with audio presentation. Both appeared in black, positioned in the center of the screen on a gray background. The images were adjusted by averaging pixel levels using Adobe Photoshop so that no differences in luminance levels occurred during the whole task.

1.5. Design

The design consisted of three within-subjects variables: Duration (30 ms, 90 ms), T2 Type (Voice, Cello, Dog), and Lag (0, 3, 9). Lag referred to the number of sounds between T1 and T2 (in Lag 0/single target conditions, no T2 was presented). The duration conditions differed not only in terms of stimulus duration but also in terms of the presentation rate (i.e., these parameters co-varied). All participants went through six blocks, each containing one of the Duration and T2 Type combinations. To minimize the block order effects, each participant was assigned to one of four versions of the experiment, where we reversed the sequential order of the block manipulations (i.e., Duration and T2 Type conditions) in half of the versions. Levels of Lag were randomly intermixed and presented equally often within each block.

The experiment contained 108 trials. Each trial included a rapid serial auditory presentation (RSAP) stream with either a single target (i.e., T1 alone) or dual-target (i.e., T1 and T2) dispersed amongst 19 or 18 distractors, respectively. A schematic representation of a dual-target RSAP stream with alternative T2 locations is shown in Fig. 1A. Each sound in a stream lasted for either 30 or 90 ms and the sounds were separated by an inter stimulus interval (ISI) of 10 ms. These durations were selected because the combined effect of stimulus duration and ISI allows us to compare whether auditory AB is more likely to occur in the typical frequency rate employed in the AB studies (10 Hz, corresponding to 90 ms stimulus presentation duration + 10 ms ISI) than at a faster rate (25 Hz, corresponding to 30 ms stimulus duration + 10 ms ISI). To keep the pupillary recording time equal, we kept the total RSAP duration the same (i.e., 2000 ms) by adding silence at the end of the stream in the 30 ms stimulus duration condition. The visual image on screen remained the same across all experimental conditions. The distracters’ order within the trials was randomized.

For half of the trials T1 was selected from the pure tones and for the rest of the times from the bell tones. T1 was always present and appeared

![Fig. 1. (A) Schematic representation of a rapid serial auditory presentation stream (RSAP) in a dual-target trial. The example depicts temporal locations of the first and the second target, with two alternative Lag positions (i.e., Lag 3 and Lag 9), amongst distracters. (B) Timeline from baseline to end of a trial. During baseline, participants were instructed to gaze at a fixation cross lasting for 500 ms with no audio presentation. This was followed by the RSAP stream presented binaurally for 2 s while participants were looking at a circle in the screen. Responses were given for T1 and T2 (in this example for human voice targets), respectively, with no time restriction. The start of the following trial was self-paced.](image-url)
at the 5th serial position in the stream. T2 varied (depending on the block) to include a Voice, Cello or Dog sound. T2 was present in 2/3 of the trials and if present, it appeared either at Lag 3 or Lag 9 (corresponding to SOA of 90 or 330 ms in the 30 ms stimulus duration condition, and 210 or 810 ms in the 90 ms stimulus duration condition).

1.6. Procedure

The experiment took place at the cognitive laboratory at the Department of Psychology, University of Oslo. The participants were tested individually in a sound-isolated, windowless room with constant luminance. After signing the consent form, participants were given verbal and written instructions about the study and were played an example of how each target sound category may sound in the context of the present study.

All participants completed six experimental blocks, each beginning with a 5-point eye calibration and its validation process. This was followed by presenting written instructions for each block, where they received detailed information about what the target categories were. Their task required reporting one or two target sounds dispersed among distracters within a rapid serial auditory presentation stream. As illustrated in Fig. 1B, before each RSAP stream, there was a 500 ms silent period where participants were requested to fixate on a cross at the center of the screen while a baseline pupil diameter was captured. After this, they were presented with the RSAP stream binaurally while keeping their gaze within a circle at the center of the screen. Following each stream, they were to identify whether the first target was a pure tone or a bell and then report whether the second target was present or absent following the first target. Responses were given using the mouse by selecting the corresponding option, as it appeared on screen, with no time restriction. No feedback was provided.

The trials were self-paced. The participants were encouraged to take a couple of minutes break and rest their eyes after completing each block. After completing the experiment, the participants received the self-report inventories of Gold-MSI with a few extra questions (i.e., de

1.7. Data processing

Data processing was carried out in R (RStudio Team, 2020). T2 performance was analyzed only for trials in which T1 was accurately identified (referred to as T2—T1 accuracy from here onward). We calculated T1 and T2—T1 accuracy for each participant and for each condition, while also applying a log-linear approach (Snodgrass and Corwin, 1988) to routinely correct for the extreme values of 0s and 1s (see Eq. 1).

\[
\text{#correct responses} + 0.5 \\
\text{#trials per condition} + 1 
\]

Eye blink artefacts were identified and omitted from the pupillary analyses using an algorithm built-in within SMI’s BeGaze software (The descriptive statistics of the eye blinks are reported in Supplementary Material). From this process, for each participant, an average of 1.2 samples (SD = 4.86) per trial were lost. The resulting gaps in the pupil data were filled using linear interpolation in R. No additional filtering or smoothing was applied. Next, we conducted a time window analysis locked to the onset of T2 in the auditory stream. For this analysis, we extracted time-locked pupillary responses from T2 onset and for a subsequent time window of 1190 ms (i.e., the remaining trial duration after T2 presentation at Lag 9 in the 90 ms condition since this is the condition where T2 appears the latest), which was kept equal among all experimental conditions. We then applied a baseline correction using a subtraction method. This involved subtracting baseline pupil response averages (i.e., 500 ms window) from the T2-locked pupil response averages. These were then used to create a data frame for each participant and for each condition.

**Statistical Analyses.** Statistical software JASP (Version 0.14.1; JASP Team, 2020) was used for statistical analyses. When the sphericity assumption was violated, a Greenhouse–Geisser correction was applied. The threshold of statistical significance was set to 0.05. A series of repeated-measures analyses of variance (rm-ANOVAs) were used to assess the impact of stimulus-related factors (i.e., Duration, T2 Type, and Lag) on performance on the auditory AB task (measured behaviorally with T1 and T2—T1 accuracy), and on LC-NA activity (measured indirectly via pupil diameter change from baseline). Single- and dual-target trials were analysed in two separate rm-ANOVAs. In dual-target trials, T2—T1 accuracy measure was based on T2 hit rates, conditional on T1 being correctly identified. In single-target trials (i.e., Lag 0), the accuracy measure reflects the correct rejection rates for T2, contingent on correct T1 identification. Correlation analyses were conducted to test the relationship between the individual-factors (i.e., musical sophistication and impulsivity), pupillary response, and AB magnitude (calculated as T2—T1 at Lag 0—T2—T1 at Lag 3).

Given that the main hypothesis is the null hypothesis (H0 or no AB for human voices), we additionally performed a Bayesian sequential t-test analysis (with default priors using JASP), given that this allows to estimate the degree of likelihood of the null hypothesis. BF01 indicates evidence in favor of the null over the alternative hypothesis. The reported BF01 value was evaluated based on the existing recommendations (Dienes, 2014; Lee and Wagenmakers, 2014), where values higher than 3 would favor the model with H0, values between 1 and 3, and 0.33 and 1 would be considered as inconclusive, and values smaller than 0.33 would favor the model with H1.

2. Results

### 2.1. Behavioural results

An overview of participants’ T1 identification performances are shown in Table 1.

**T2—T1 Accuracy.** As our analysis of interest concerns the attentional blink, we focus our analyses on the dual-target trials alone. The results obtained from the single-target trials and T1 accuracy can be found in Supplementary Material.

A rm-ANOVA revealed significant main effects of T2 Type \(F(2, 110) = 7.73, p < 0.001, \eta^2_p = 0.12\) and Lag \(F(1, 55) = 10.66, p = 0.002, \eta^2_p = 0.16\) on T2—T1 accuracy. Most importantly, the three-way interaction of T2 Type × Duration × Lag was statistically significant, \(F(2, 110) = 3.89, p = 0.02, \eta^2_p = 0.07\). Fig. 2 depicts the difference between T2—T1 accuracy rates at Lag 3 and Lag 9, as a function of T2 Type, separately for each stimulus duration. Planned contrasts revealed that in the 30 ms condition, none of the T2 Types differed between Lag 3 and Lag 9 \(t = -1.24, p = 0.25\) for cello; \(t = 2.85e-15, p = 1.00\) for dog; \(t = -1.31, p = 0.19\) for human voice T2s). However, in the 90 ms condition, there was

**Table 1**

| T1 accuracy (Mean and SE) as a function of T2 presence, Lag, T2 Type and Stimuli Duration. |
|-----------------------------------------------|-----------------------------------------------|
| T2 Type and Duration | Absent Lag 0 | T2 Presence | Lag 3 | Present | Lag 9 |
|----------------------|--------------|-------------|-------|---------|-------|
| Human voice | | | | |
| 30 ms | 0.83 (0.03) | 0.92 (0.04) | 0.33 (0.03) | 0.82 (0.03) |
| 90 ms | 0.90 (0.04) | 0.98 (0.04) | 0.97 (0.03) | 0.97 (0.03) |
| Cello | | | | |
| 30 ms | 0.90 (0.04) | 0.97 (0.03) | 0.86 (0.03) | 0.97 (0.03) |
| 90 ms | 0.97 (0.04) | 0.75 (0.05) | 0.97 (0.03) | 0.97 (0.03) |
| Dog | | | | |
| 30 ms | 0.89 (0.04) | 0.90 (0.03) | 0.84 (0.04) | 0.84 (0.04) |
| 90 ms | 0.75 (0.04) | 0.69 (0.04) | 0.88 (0.04) | 0.88 (0.04) |
lower accuracy at Lag 3 than at Lag 9 for cello \( t = -3.68, p < 0.001 \) and dog \( t = -2.45, p = 0.01 \) T2s, while – as expected – there was no significant difference for human voice T2s \( t = 0.79, p = 0.43 \).

Regarding the main effects, Bonferroni-adjusted post hoc comparisons of T2 Type showed that when T2 was a dog sound, T2—T1 accuracy was significantly worse than for a cello sound (Mean difference = -0.10, SE = 0.03, \( p = 0.001 \)) or human voice (Mean difference = -0.09, SE = 0.03, \( p = 0.006 \)). Planned contrasts with Lag yielded a significant difference in T2—T1 accuracy between the two lags \( t(55) = -0.07, p = 0.002 \), with lower accuracy rate at Lag 3 than Lag 9 (Mean difference = -0.07, SE = 0.02), reflecting the presence of an overall auditory AB effect.

T2—T1 accuracy was, however, affected by Duration \( F(2, 110) = 3.45, p = 0.03, \eta^2_p = 0.06 \). Post hoc tests revealed that while T2—T1 accuracy did not differ significantly for different T2 categories in the 30 ms condition (all \( p > 0.05 \)), in the 90 ms condition the accuracy was significantly lower for dog than for both cello (Mean difference = -0.12, SE = 0.04, \( p = 0.01 \)) and human voice T2s (Mean difference = -0.15, SE = 0.04, \( p < 0.001 \)).

None of the other interactions with Lag reached statistical significance (T2 Type × Lag: \( p = 0.12 \), Greenhouse-Geisser corrected; Duration × Lag: \( p = 0.28 \)).

We predicted that the human voices would be the least likely to show an auditory AB effect (essentially a ‘null’ effect). This was supported by non-significant differences in accuracy in the 90 ms condition across Lags in the ANOVA. However, we also conducted a sequential \( t \)-test Bayesian analysis to further confirm our main but negative hypothesis. Fig. 3 shows not only that the null hypothesis was supported throughout the sequential development of the data for this condition, but also that the evidence in support of the null hypothesis (H0) was ‘strong’ in our sample for human voice T2s in the 90 ms condition \( (BF_0- = 11.02) \), thus conclusively supporting the absence of an AB effect for human voice T2s in the 90 ms condition.

### 2.2. Pupillary results

We examined the influence of stimulus-related factors on participants’ pupillary response (i.e., T2-time-locked changes in pupil diameter) under each condition during an AB task. A \( 2 \times 2 \times 3 \) rm-ANOVA on
the pupil data yielded a significant main effect of Lag \( F(1,55) = 53.13, p < 0.001, \eta^2_p = 0.49 \) as well as a two-way interaction of Duration \( \times \) Lag \( F(1,55) = 35.87, p < 0.001, \eta^2_p = 0.39 \) and a three-way interaction of T2 Type \( \times \) Duration \( \times \) Lag \( F(2,110) = 4.76, p = 0.01, \eta^2_p = 0.08 \). The main effects of T2 Type \( (p = 0.25) \) and Duration \( (p = 0.21) \) as well as the interaction effects of T2 Type \( \times \) Duration \( (p = 0.26) \) and T2 Type \( \times \) Lag \( (p = 0.08, \text{Greenhouse-Geisser corrected}) \) were not statistically significant.

Fig. 4 shows T2-time-locked pupil responses at Lag 3 and Lag 9 for each T2 Type, separately for the two stimulus duration conditions, and target type. There was an increase in the pupillary responses when T2 followed T1 at Lag 9 compared to Lag 3, but only in the 90 ms condition. Planned contrasts revealed that in the 30 ms condition, T2-time-locked pupil responses did not differ across the two lag conditions for any of the T2 categories (cello: \( t(323.85) = 0.88, p = 0.38 \); dog: \( t(323.85) = 1.64, p = 0.10 \); human voice: \( t(323.85) = -0.28, p = 0.78 \)). In the 90 ms condition, however, statistically significant differences between T2-time-locked pupil responses in the two lag conditions were observed for all three T2 categories (cello: \( t(323.85) = 2.67, p = 0.01 \); dog: \( t(323.85) = 6.76, p < 0.001 \); human voice: \( t(323.85) = 7.10, p < 0.001 \)). That is, participants’ pupil dilations tended to be larger in the longer stimulus duration and longer inter-item lag conditions compared with the shorter duration and shorter lag conditions.

### 2.3. Questionnaire results

**Musical Sophistication.** General musical sophistication scores were on average 75.30 ± 23.27 SD, which ranked at the 36th percentile based on the data norms (Müllensiefen et al., 2014). We expected negative correlations between the musical sophistication scores and the auditory AB magnitude, and with the pupil averages. To this aim, we calculated AB magnitude and pupil dilation averages for each participant. Contrary to our expectation, none of the Gold-MSI factors (including the general factor of musical sophistication) correlated with the attentional blink magnitude or with the pupil averages in 30 and 90 ms. The detailed results are shown in Supplementary Table B4.

**Impulsivity.** The average score was 2.09 ± 2.51 on dysfunctional and 5.34 ± 2.20 on functional impulsivity component. Here we predicted that functional impulsivity scores would be negatively associated with both the AB magnitude and the pupil averages. However, neither functional nor dysfunctional impulsivity correlated with the attentional blink magnitude or with the pupillary response averages in the 30 and 90 ms conditions at 0.05 significance threshold (see Supplementary Table B4).

### 3. Discussion

The present findings confirm the occurrence of an AB ‘attenuation’ effect with the human voice when it is the second target in an auditory temporal selective attention task. That is, we observed no reliable auditory AB effect for human voice T2s since the accuracy of target detections remained the same in the 90 ms condition. Such a finding is also consistent with the main finding reported by Akça et al. (2020) using 150 ms stimulus duration. In addition, we found that the AB effect was more likely to occur for cello sounds than for dog sounds, which suggest that biological and/or familiar sounds, rather than acoustical similarity (e.g., between the voice and the cello) play a crucial role in AB strength.

However, the duration of individual sounds was relevant in determining the range under which an AB was observed. Our findings indicate that the auditory AB effect was more likely to occur under 90 than 30 ms conditions. Considering presentation rates of information in the rapid auditory streams, we found behavioral evidence for an auditory AB at a relatively slower presentation rate of 10 Hz, but not at the faster rate of 25 Hz. At first sight, this finding seems to diverge from previous auditory AB studies (Arnell and Jolicoeur, 1999; Shen and Alain, 2010) that suggest auditory AB is more likely under faster than slower presentation rates. One reason for the discrepancy between our results and those of previous auditory AB studies may lie in what constitutes “slow” and “fast” in those studies versus the present study. For example, in Arnell and Jolicoeur (1999) auditory AB was found in 105 and 120 ms rate but not in 135 and 150 ms/item rate. Similarly, in Shen and Alain (2010), the fastest rate was 90 ms/item. Hence, what would be considered as the “fastest” rate in those studies is more or less similar to the “slower” of the two presentation rates in the present study.

Moreover, Shapiro and colleagues (Shapiro et al., 2017) suggest that the presentation frequency range of the stimulus presentation stream is the most critical factor in determining the AB. Their study demonstrates how the (visual) AB rises and falls as a function of the frequency range, and most critically, that the AB might be restricted to 10 and 15 Hz frequency range. In particular, the (visual) AB effect was significantly reduced in both slower (i.e., around 6 Hz) and faster (i.e., around 36 Hz) presentation frequencies. Drawing from the electrophysiological studies on brain oscillations (e.g., Dijk et al., 2008), it has been argued that the AB might be mediated by alpha/beta frequencies, as increased alpha and beta oscillations constitute the most unfavorable condition for perception (Shapiro et al., 2017; Shapiro and Hanslmayr, 2014). Given that a large body of the existing AB studies are reported in this sweet spot of 10 to 15 Hz, we believe it is important to compare frequency ranges within and outside of this ‘sweet spot’ to better account for the effects of stimulus-factors on auditory AB. This was another motivation for the
chosen frequencies (and durations) in the present study. Our findings are consistent with the idea that the AB phenomenon might be restricted to certain presentation frequencies, as previously found in the visual domain (Shapiro et al., 2017). Additional empirical evidence is needed to confirm that the auditory AB effect is confined to alpha and beta frequencies.

Another possible explanation for why auditory AB effect was more likely to occur under 90 than 30 ms condition could be related to backward masking. In the auditory AB literature, the findings of Vachon and Tremblay (2005) suggest that, just as in visual AB effects, backward masking of T2 (i.e., presenting distractors after T2) is essential for observing auditory AB effects. Based on this, one could argue that the absence of AB in the 30 ms condition could potentially be explained by insufficient backward masking (i.e., if brief information sequences can be more easily retained in echoic memory). The absence of AB for voices in the 90 ms condition, however, is unlikely to be explained by the masking requirement. This is because auditory AB was present for cello and dog T2s, even though the masking interference remained the same across all T2 conditions.

This was the first study to explore the auditory AB effect with pupillometry. We used the pupil dilation response as an index of the attentional capture of the auditory stimuli. In the field of music cognition, changes in pupil size are used as a reliable and valid measure to index allocation of attentional resources during music listening or performance (e.g., Skaansar et al., 2019; Endestad et al., 2020; Bishop et al., 2021; Laeng et al., 2021; Speich et al., 2022). In the present context, we revealed fluctuations in attentional deployment during the auditory AB task for the different T2 types. The behavioral findings were generally mirrored by the fluctuations in pupillary dilations during the task. Time-locked pupillary analyses to the onset of T2 revealed dramatic increases in the allocation of attention at Lag 9 compared to Lag 3 only in the 90 ms condition. That is, during an AB (Lag 3 in the present case), the pupils remained relatively small indicating lower attentional capture of the targets.

In other words, the pupil diameter was larger in the long lag condition, where an AB typically recovers. Assuming changes in pupil dilation reflect the “intensity” of attention (Kahneman, 1973), this would mean that the amount of cognitive processing allocated to target detection was greater when T2 was presented at Lag 9. Alternatively, this finding can also reflect better attentional allocation in the long compared to the short lag condition. Given that T1 accuracy was also reduced at short lags for dog and cello targets in 90 ms, the auditory AB observed in these conditions is not attributable to participants adopting a particular strategy, but instead could be a result of inefficient deployment of attention to T2 while processing T1. Supporting this interpretation, numerous EEG studies have shown that the N2pc, an ERP component which is indicative of attentional capture of task-relevant stimuli among distractors, is elicited in the long lag trials, but substantially attenuated in short lag trials during an AB task (for a review, see Martens and Wyble, 2010). Generally, the present findings are also consistent with previous research (Wierda et al., 2012; Zylberberg et al., 2012) in showing that pupil dilation can index modulations of LC-NA activity and reveal the load on cognitive processing or attentional allocation within an AB task. However, to our knowledge, this is the first study to relate the modulations of LC-NA system’s activity to auditory AB effects.

The pattern of results obtained with behavioural measure of AB and pupil responses was not identical. In particular, in the 90 ms condition, we see statistical evidence that voices escape the AB only in the behavioural accuracy, but not in pupil responses. We think that this is not a failure of two methods to converge; rather, the behavioural accuracy provides us evidence of voices escaping the AB, while pupillary changes tell us about the mechanism behind ‘escaping the blink’. As reported by Zylberberg et al. (2012), the reduced pupil dilation for shorter lags, in which the AB typically occurs, can be a trace of interference even without a behavioral manifestation. Indeed, we observed smaller pupil sizes at the shorter Lag...
result of presenting more sounds). Future studies could benefit from more systematic testing of the effects of these methodological choices. Moreover, it may be of interest to include additional measures such as reaction time and certainty ratings to better elucidate the underlying cognitive processes in those cases where there was no behavioral manifestation of an AB effect. Finally, contrary to our hypothesis, our data did not yield any significant correlations with the musicality or impulsivity scores. Whether these or other participant factors contribute to the auditory AB effects and/or the amount of mental effort exerted in the detection of auditory targets remains an open question.

4. Conclusion

All in all, our results support the conclusion that human voices escape the AB and that pupillary changes, modulated by inter-target Lag and T2 type, are generally consistent with the so-called T2 attentional deficit which is behaviorally observed in the auditory attentional blink task. In addition, we found support for the hypothesis that human voices could require a less intense allocation of attention, or noradrenergic potentiation, compared to other auditory stimuli.

5. Ethics Statement

This study was performed in accordance with the recommendations of the Declaration of Helsinki. A written informed consent was obtained from all participants. The study protocol was approved by the Norwegian Centre for Research Data (NSD) on 13 February 2018 [Ref. no: 58784/3].

6. Author Contributions

MA and BL conceptualized the study; MA, LB, and BL developed the study design and methodology; MA prepared the stimuli and programmed the experiment; LB curated the data; MA performed the formal statistical analyses with supervision of LB, BL, and JKV; MA wrote the first draft of the manuscript; LB, BL, and JKV supervised the project; BL, LB, and JKV provided critical commentary and contributed to manuscript revisions. All authors read and approved the submitted version.

7. Funding

This work was partially supported by the Research Council of Norway through its Centres of Excellence scheme, project number 262762.

8. Data Availability Statement

The datasets generated for this study can be found in the Supplementary Material.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Sumeyya Atmaca for her assistance in participant recruitment and data collection.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.bandc.2022.105928.

References

Agus, T. R., Suirod, C., Thorpe, S. J., & Pressnitzer, D. (2012). Fast recognition of musical sounds based on timbre. The Journal of the Acoustical Society of America, 131(5), 4124–4133.
Akç, M., Laeng, B., & Gody, R. I. (2020). No evidence for an auditory attentional blink for voices regardless of musical expertise. Frontiers in Psychology, 10, 2935.
Alonzo, D., Sneve, M. H., Espeth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. Journal of Vision, 14(4), 1–20.
Arnemo, M. J., & Jolicoeur, P. (1999). The attentional blink across stimulus modalities: Evidence for central processing limitations. Journal of Experimental Psychology: Human Perception and Performance, 25(3), 630.
Askenfelt, A. (1991). Voices and strings: Close cousins or not? In J. Sundberg, L. Nord, & R. Carlson (Eds.), Music, language, speech and brain (pp. 243–256). Palgrave.
Aw, E., flerencen, J., Laurey, P., Dalhval, H., van der Jagt, T., & Dassonville, P. (2004). Evidence against a central bottleneck during the attentional blink: Multiple channels for configurational and feature processing. Cognitive Psychology, 48(1), 95–126.
Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin, 91(2), 276.
Belin, P. (2017). Similarities in face and voice cerebral processing. Visual Cognition, 25(4–6), 658–665.
Bishop, L., Jensenius, A. R., & Laeng, B. (2021). Musical and bodily predictors of mental effort in street quartet music: An ecological pupillometry study of performers and listeners. Frontiers in Psychology, 2220.
Boulanger, R. (2007). The berklee college of music sampling archive (vol. 5).
Bruno, N., Roux, S., Clery, H., Rogier, O., Bidet-Caulet, A., & Barthélémy, C. (2013). Early neurophysiological correlates of vocal versus non-vocal sound processing in adults. Brain Research, 1528, 20–27.
Brydges, B. (2019). How many participants do we have to include in properly powered experiments? a tutorial of power analysis with reference tables. Journal of Cognition, 2(1), 16.
Charest, I., Pernet, C. R., Rousselet, G. A., Quíones, I., Latium, M., Sarah, F.-B., Charron, J.-P., & Belin, P. (2009). Electrophysiological evidence for an early processing of human voices. BMC Neuroscience, 10(1), 1–11.
Dickman, S. J. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. Journal of Personality and Social Psychology, 58(1), 95.
Dienes, Z. (2014). Using bayes to get the most out of non-significant results. Behavioral Research Methods, 46(2), 136–147.
Dijk, H. V., Schoffelen, J.-M., Oostenveld, R., & Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. Journal of Neuroscience, 28(8), 1816–1823.
Endestad, T., Gody, R. I., Sneve, M. H., Hagen, T., Bochymka, A., & Laeng, B. (2020). Mental effort when playing, listening, and imagining music in one pianist’s eyes and brain. Frontiers in Human Neuroscience, 14, 576086.
Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. Current Opinion in Neurobiology, 11(2), 219–224.
Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. Science, 143(3611), 1190–1192.
Jackson, M. C., & Raymond, J. E. (2006). The role of attention and familiarity in face identification. Perception & Psychophysics, 68(4), 543–557.
JASP Team. (2020). Jasp (version 0.14.1) [computer software]. https://jasp-stats.org/.
Josh, S., & Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. Trends in Cognitive Sciences, 24(6), 466–480.
Josh, S., & Gold, J. I. (2022). Context-dependent relationships between locus coeruleus firing patterns and coordinated neural activity in the anterior cingulate cortex. Elife, 11, e65490.
Kahnmann, D. (1973). Attention and effort (Vol. 1063). Englewood Cliffs, NJ: Prentice Hall.
Kahnmann, D., & Beatty, J. (1966). Pupil diameter and load on memory. Science, 154(3756), 1583–1585.
Laeng, B., Garvisl, L., Leseth, G., Elkemo, M., Ernst, G., & Leknes, S. (2021). ‘defrosting’ music chills with naltrexone: The role of endogenous opioids for the intensity of musical pleasure. Consciousness and Cognition, 90, 103102.
Landau, A. M., & Bentin, S. (2008). Attentional and perceptual factors affecting the attentional blink for faces and objects. Journal of Experimental Psychology: Human Perception and Performance, 34(4), 818.
Lee, M. D., & Wagenmakers, E.-J. (2014). Bayesian cognitive modeling: A practical course. Cambridge University Press.
Levy, D. A., Granot, R., & Bentin, S. (2003). Neural sensitivity to human voices: ERP evidence of task and attentional influences. Psychophysiology, 40(40), 291–305.
Li, C. R., Chen, S., Lin, W., & Yang, Y. (2005). Attentional blink in adolescents with varying levels of impulsivity. Journal of Psychiatric Research, 39(2), 197–205.
Liu, Y., Rodenkirch, C., Moskowitz, N., Schriver, B., & Wang, Q. (2017). Dynamic lateralization of pupil dilation evoked by locus coeruleus activation results from sympathetic, not parasympathetic, contributions. Cell Report, 20(13), 3099–3112.
Martens, S., & Johnson, A. (2006). Working memory capacity, intelligence, and the magnitude of the attentional blink revisited. Experimental Brain Research, 192(1), 43–52.
Martens, S., Munneke, J., Smid, H., & Johnson, A. (2006). Quick minds don’t blink: Electrophysiological correlates of individual differences in attentional selection. Journal of Cognitive Neuroscience, 18(9), 1423–1438.
Martens, S., & Valchev, N. (2009). Individual differences in the attentional blink: The important role of irrelevant information. Experimental Psychology, 56(1), 18–26.
