Running Head: Surprise-induced Deafness

Surprise-induced Deafness: Unexpected Auditory Stimuli Capture Attention to the Detriment of Subsequent Detection

Takashi Obana[123], Stephen Wee Hun Lim[2], & Christopher L. Asplund[1234]

1 N.1 Institute for Health (formerly SINAPSE), National University of Singapore, Singapore

2 Department of Psychology, National University of Singapore, Singapore

3 Division of Social Sciences, Yale-NUS College, Singapore

4 Clinical Imaging Research Centre, Yong Loo Lin School of Medicine, Singapore

Author Note: This work was supported by a Heads and Deanery Research Support Scheme grant to S.W.H.L. It was also supported by a Singapore Ministry of Education and Yale-NUS College start-up grant, as well as a Singapore Ministry of Defence and SINAPSE grant, both to C.L.A. The authors would like to thank Wang Yushi and Lee Sze Chi for their assistance with data collection, Helen Jin for stimulus preparation, and Garry Kong and Daryl Fougnie for helpful discussion.

Address for correspondence: Christopher L. Asplund, Yale-NUS College, 28 College Avenue West #01-501, Singapore 138533, chris.asplund@yale-nus.edu.sg

Abstract

Our attention is often captured by unexpected or unusual sounds. Such stimulus-driven control of attention can be adaptive, as potentially relevant events need to be quickly evaluated and acted upon. Attentional capture, however, comes with a cost: Ongoing tasks may be disrupted. In a series of seven experiments (n=773), we investigated the effects of task-irrelevant, rare, and relatively unexpected sounds ("surprise stimuli") on probe detection in rapid auditory presentation (RAP) streams. Surprise stimuli caused "Surprise-induced Deafness" (SiD), a severe detection deficit that lasted for under one second within each trial and gradually habituated across several trials. SiD was sensitive to informational "surprise", with larger deficits following stimuli that were infrequent or varied across trials. The effect also generalized: Natural sounds or constructed stimuli could disrupt detection of either spoken letters or simple tones. We also compared SiD to the auditory attentional blink (AAB), a similar paradigm in which goal-directed target processing disrupts probe detection. We found that the two deficits were weakly correlated. We conclude that SiD is a novel perceptual deficit that primarily reflects stimulus-driven attentional capture. It may involve other forms of attentional control as well, thereby reflecting multiple attentional influences on awareness.

Key phrases: auditory attention, surprise, stimulus-driven attention, individual differences, attentional capture

General importance

When an unexpected sound occurs, it often grabs our attention. Our focus shifts to the sound so that we can assess its importance. The sound could be a warning of imminent danger; alternatively, it could be someone's innocuous ringtone. In this study, we show that there is a cost when a sound grabs our attention. Specifically, we may fail to hear a subsequent sound for which we were listening, even when the unexpected and expected sounds come from the same source. This effect, which we term “Surprise-induced Deafness”, lasts for less than a second after each unexpected sound, and it diminishes as the novelty of such sounds wears off. Our findings could be relevant to the design of user interfaces. More broadly, they improve our theoretical understanding of the trade-off between staying on task and paying attention to what is new.

Introduction

An organism's survival depends on its successful detection of unexpected, sudden changes in its surroundings. Hearing is especially important for this function because it can serve as an "early warning system" (Scharf, 1998), sensitive to changes regardless of their location relative to gaze and body position. In addition to these sensory features, the human auditory system includes cognitive processes to rapidly assess sounds for novelty or deviance from expectation (Fritz, Elhilali, David, & Shamma, 2007). A stimulus that warrants further evaluation may automatically become the focus of selective attention, thereby aiding perception and the production of an appropriate behavioral response. In addition to attention's faciliatory roles, however, it has a dark side: Unattended items often fail to reach awareness (Chun & Marois, 2002). In this study, we tested whether non-spatial attentional capture could produce such profound deficits in auditory awareness. We also sought to characterize the factors that influence whether attention is redirected in the first place, as such factors could further elucidate the mechanisms by which attention is controlled.

The auditory system, like other sensory systems, is sensitive to the salience of environmental stimuli. Many factors can render a stimulus salient, including behavioral relevance in a given context, stimulus intensity, frequency of occurrence, novelty, and deviance from expectation (Downar, Crawley, Mikulis, & Davis, 2002; Fritz et al., 2007). These latter three factors involve the idea of informational "surprise", as the event carries information by violating ongoing expectations about the world (Asplund, Todd, Snyder, Gilbert, & Marois, 2010; Horstmann, 2015; Meyer, Niepel, Rudolph, & Schützwohl, 1991; Niepel, Rudolph, Schützwohl, & Meyer, 1994). The brain is sensitive to such "surprises". For example, novel or deviant sounds evoke an electroencephalographic and magnetoencephalographic response termed the mismatch negativity (MMN) (Escera, Alho, Winkler, & Näätänen, 1998; Friedman, Cycowicz, & Gaeta, 2001; Näätänen & Alho, 1995;

Näätänen, Paavilainen, Rinne, & Alho, 2007; Niepel et al., 1994; Parmentier, 2014). The MMN emerges rapidly following stimulus presentation (100-200 ms), after which the evoking stimulus may become attended if further investigation is required (Friedman et al., 2001; Horváth, Winkler, & Bendixen, 2008). In the case of a novel stimulus, a subsequently evoked component (the novelty-P3) is thought to index an attentional shift towards the stimulus (Courchesne, Hillyard, & Galambos, 1975; Friedman et al., 2001; Friedman & Simpson, 1994).

Behaviorally, the response to a relatively unexpected or novel stimulus is often orientation towards that stimulus. This “orienting response” (OR) includes positioning the appropriate receptor organ to investigate a potentially relevant stimulus or change in the environment (Pavlov, 1927; Sokolov, 1990). In addition to overt physical reorientation, such as turning one’s head to better localize a sound, the OR involves cognitive reorientation: Attentional resources are reallocated in order to process the stimulus or event (Downar et al., 2002; Escera et al., 1998; Kahneman, 1973). The novelty-P3 may represent this cognitive aspect of the OR, which involves a spatial or non-spatial shift of attention (Escera, Alho, Schröger, & Winkler, 2000; Friedman et al., 2001; Knight & Scabini, 1998). Importantly, both the novelty-P3 and the behavioral components of the OR rapidly habituate as the novelty of the evoking stimulus declines, such as following repeated presentations.

Although the redirection of attention to an OR-evoking or otherwise distracting stimulus enhances the processing of that stimulus, there are also negative effects of such attentional shifts. In particular, ongoing task performance may suffer. Performance impairments on a primary task are often used to study auditory distraction and attentional capture. For example, in a task where participants had to quickly report which of two speakers produced a target tone, responses were slowed by the presentation of a novel and task-irrelevant sound (Niepel et al., 1994). Indeed, the presentation of deviant sounds generally slows target detection and discrimination (Dalton & Lavie, 2004; Horváth & Winkler, 2010; Roeber, Berti, & Schröger, 2003; Schröger, 1996; Vachon, Labonté, & Marsh, 2017). In some cases, primary task accuracy also suffers (Horváth & Winkler, 2010; Schröger, 1996).

Speed and accuracy impairments are often contingent on the time between the distracting stimulus and the target item. Most impairments follow short (125-200 ms) SOAs (stimulus onset asynchronies) (Horváth & Winkler, 2010; Niepel et al., 1994; Schröger, 1996; Vachon et al., 2017), and no impairments were found following 560 ms (Schröger, 1996) or 1500 ms (Niepel et al., 1994) SOAs (though see Roeber et al., 2003). The timecourse of attentional capture caused by “surprising” stimuli is grossly similar in the visual domain, though with an important exception: Although attentional capture is observed with short SOAs (100-200 ms), it tends to be more robust for SOAs of several hundred milliseconds (Asplund, Todd, Snyder, Gilbert, et al., 2010; Horstmann, 2005, 2015). In Surprise-induced Blindness, for example, a relatively unexpected and task-irrelevant stimulus was found to impair detection of a subsequent probe in a rapid serial visual presentation (RSVP) stream of distractors (Asplund, Todd, Snyder, Gilbert, et al., 2010). Although a 100-150 ms SOAs yielded a mild impairment, a far more severe deficit was evidenced for 350-400 ms SOAs (with no deficit following a 780 ms SOA). Furthermore, only this severe deficit rapidly habituated after 1-6 presentations of surprising stimuli, akin to the habituation profiles for both the orienting response (Kahneman, 1973; Pavlov, 1927; Sokolov, 1990) and the novelty-P3 (Escera et al., 2000; Friedman et al., 2001; Knight & Scabini, 1998). It is unclear whether auditory attentional capture under similar experimental conditions would show such timecourses; to the best of our knowledge, a comprehensive test of auditory attentional capture across various SOAs has yet to be conducted. Comparable timecourses would imply similar mechanisms of attentional control for each sensory modality, or even supramodal responses to surprising events (Horstmann, 2015). Conversely, distinct timecourses would suggest distinct mechanisms.

One way to better understand auditory attentional capture, as well as the similarity to its visual counterpart, would be to construct and test an analogue of Surprise-induced Blindness (i.e. Surprise-induced Deafness). Although such a paradigm would test attentional capture, it would also be procedurally and conceptually similar to the auditory attentional blink (AAB). In the AAB, participants search for two auditory targets in a stream of distractor sounds; the second target is often missed when it closely follows the first. In many AAB studies, performance increases monotonically with increasing SOA (Horváth & Burgyán, 2011; Mondor, 1998; Potter & Chun, 1998; Shen & Mondor, 2006; Vachon & Tremblay, 2005). In other AAB studies, however, the most severe deficits occur with SOAs around 300-400 ms, at least under some experimental conditions (Arnell & Jolicoeur, 1999; Martens, Johnson, Bolle, & Borst, 2009; Tremblay, Vachon, & Jones, 2005). The timecourse of the AAB may therefore be similar to both visual and auditory attentional capture effects, perhaps indicating similar causes.

Despite the similarities between the auditory attentional blink (AAB) and paradigms that reflect attentional capture effects—potentially including Surprise-induced Deafness—, they critically differ in how attention is allocated to the deficit-inducing stimulus. Traditionally, attention has been thought to be controlled in either a top-down, goal-directed manner or a bottom-up, stimulus-driven one (Debener, Kranczioch, Herrmann, & Engel, 2002; Egeth & Yantis, 1997). The AAB appears to involve primarily the former, as the deficit is absent or greatly attenuated when the first target item is ignored. Conversely, attentional capture is often ascribed to the latter form of attentional control. There has been longstanding debate, however, over whether attentional capture is due exclusively to stimulus-driven factors that are independent of behavioral goals. An alternative explanation for attentional capture is given by the contingent capture hypothesis, which posits that stimuli capture attention only if they contain features that define targets (Folk, Remington, & Johnston, 1992). Recently, both contingent and non-contingent capture have been demonstrated in the auditory domain (Vachon et al., 2017). Although Asplund et al. (2010) argued that Surprise-induced Blindness involved only non-contingent capture, Surprise-induced Deafness could involve either or both forms depending on the task procedures and differences between the visual and auditory systems.

Another possibility is that the dichotomy of attentional control itself is incomplete (Awh, Belopolsky, & Theeuwes, 2012). That is, other sources of control must be considered in order to explain the range of behavioral results. For example, the history of responses to a given stimulus biases attention in a way that may be incongruent with either current task goals or the bottom-up properties of the stimulus (Belopolsky, Schreij, & Theeuwes, 2010). Different aspects of salience may fall into this additional category of attentional control as well. For example, the rapid habituation of attentional capture effects implies that a dynamically-updated internal model influences whether a stimulus captures attention. Such capture may nevertheless go against the individual’s task goals and voluntary control of attention.

The present experiments

In this study, we describe a series of experiments designed to investigate auditory attentional capture by “surprising” stimuli. In an initial experiment, we attempted to establish and characterize Surprise-induced Deafness, an auditory analogue of Surprise-induced Blindness in which a relatively rare and unexpected stimulus causes a subsequent probe item to be missed. We also sought to test the temporal properties of any such deficit, both within each trial and across multiple trials containing surprising stimuli. We expected pronounced deficits for SOAs around 300-400 ms, with smaller effects for shorter SOAs (~100 ms) and no effects for far longer ones (Asplund, Todd, Snyder, Gilbert, et al., 2010; Horváth & Burgyán, 2011; Meyer et al., 1991; Niepel et al., 1994; Schröger, 1996). We also predicted that SiD effects, particularly those for intermediate SOAs, would habituate after repeated presentations of surprising items (Asplund, Todd, Snyder, Gilbert, et al., 2010; Friedman et al., 2001; Kahneman, 1973; Pavlov, 1927; Sokolov, 2002). Given that auditory novelty-P3s largely attenuate after approximately six stimulus exposures, we expected the habituation rate to be comparable in the auditory and visual domains (Friedman & Simpson, 1994).

In six subsequent experiments, we further explored the SiD phenomenon and its relevance to attentional control. Foremost, it was necessary to simplify the paradigm so that it could be run online using Amazon Mechanical Turk (AMT; Experiment 2). We then tested which aspects of salience modulate or are necessary for SiD. Specifically, we examined the effects of the surprising stimuli's relative novelty and frequency of appearance (Experiment 3), as well as the effects of their gross behavioral relevance and low-level stimulus properties (Experiment 4). We also considered whether participants adopted search strategies that may have introduced more goal-directed attentional elements into the task, as opposed to primarily stimulus-driven components (Experiment 5). Finally, we conducted a correlation experiment to test the relationship between Surprise-induced Deafness and the auditory attentional blink (Experiments 6 and 7).

Experiment 1: Establishing and characterizing Surprise-induced Deafness (SiD)

In this first experiment, we tested whether the presentation of an unexpected and task-irrelevant auditory stimulus (the "surprise" stimulus) would impair the detection of a subsequent probe item. We also sought to characterize any such deficit's timecourse, both the expected transience within each trial and the expected habituation across them. Although our primary dependent measure was detection rate, we also recorded and analyzed reaction times because auditory masking may be ineffective (Crowder, 1993; Potter & Chun, 1998). Finally, we tested for the generalizability of any observed deficits by swapping the stimulus categories of distractors and probes as compared to surprise stimuli (Experiments 1A and 1B). These versions of the experiment are presented in sequence, followed by a comparison between them.

To test each of the ideas above, we adapted the Surprise-induced Blindness paradigm (Asplund, Todd, Snyder, Gilbert, et al., 2010) from the visual domain for the auditory domain. We predicted a similar pattern of results for Surprise-induced Deafness (SiD), namely maximal deficits from intermediate SOAs (300-400 ms), rapid habituation of the effects from such SOAs, and the presence of SiD regardless of the particular stimulus categories used.

Experiment 1A: SiD in a tone detection task

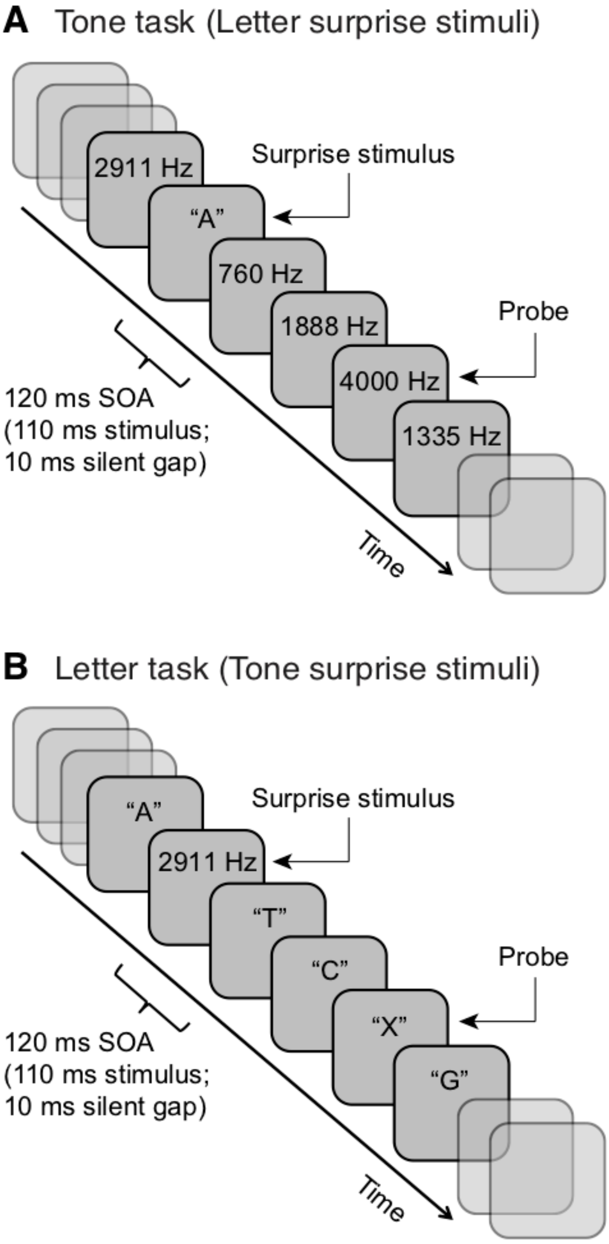
Method

Participants. Forty-nine individuals from the National University of Singapore (NUS) community (20 males, 4 gender not reported, mean age = 22.38 years, age range = 19-27) with self-reported normal hearing participated for either payment (S$5) or participation credit in the NUS Department of Psychology research pool. Attentional capture effects from unexpected stimuli are often large (e.g. 0% detection following first Surprise stimulus versus 81% baseline detection for Surprise-induced Blindness (Asplund, Todd, Snyder, Gilbert, et al., 2010)), but we recruited a larger sample here in case auditory effects were smaller and to confidently assess effect habituation. The NUS Institutional Review Board approved of the protocol for this experiment and all subsequent experiments reported here.

Stimuli and apparatus. The probe stimulus was a 4000 Hz pure tone, and 21 other pure tones (log-related frequencies between 639 and 3175 Hz) served as distractors. Unexpected stimuli ("surprises") were spoken letters of the alphabet, excluding W, N, F, S, and X. The letters, spoken by a female native English speaker, were recorded using an

Olympus IC recorder and Bose microphone. All sounds were compressed to a duration of 110 ms (120 ms for practice trials) without changing the pitch in Audacity (The Audacity Team). All sound stimuli were then adjusted so that their maximum amplitudes were equivalent. Based on reports from pilot participants and the experimenters, the tones and spoken letters were perceived to have roughly equivalent volumes when the former were reduced to 15% of the latter's amplitude; the stimuli were adjusted accordingly. PsychoPy software (Peirce, 2007) and a Dell computer (OPTIPLEX 990) were used to control stimulus presentation and for data logging. Sounds were presented binaurally through TDK headphones (ST 100) with an approximate maximum volume of 70 dB. Participants made their responses through a standard computer keyboard.

Task procedures. Each participant-initiated trial began with the presentation of the probe tone to refresh the participant's memory. A white fixation cross was displayed in the center of the screen until participants pressed any key to commence each trial. A rapid auditory presentation (RAP) of 30 tones then began, with each tone sounding for 110 ms with a 10 ms intervening gap (Figure 1A). The onset and offset of each sound included a 2-ms linear amplitude ramp to eliminate clicking or popping of the stream. During 75% of the trials, the probe was presented after 17 to 27 RAP items. Participants were instructed to press the '1' key in response to the probe as quickly as possible without sacrificing accuracy. If no response was made during the RAP, participants were prompted by the question "Target?" 300 ms after the end of the RAP. (Note that we referred to the probe as a "target" with our participants. We use the term "probe" in our descriptions here for clarity and consistency with Experiments 6 and 7.) Participants pressed the '1' key if they judged that the probe had been presented and the '0' key if not.



A Tone task (Letter surprise stimuli)

B Letter task (Tone surprise stimuli)

Figure 1. The sequence of stimuli in a rapid auditory presentation (RAP) stream in (A) Experiment 1A (tone task) and (B) Experiment 1B (letter task). A Lag 3 (360 ms SOA) surprise trial is shown for each.

During 10% of the trials, a "surprise" stimulus replaced a distractor stimulus with a surprise-probe lag of 1, 3, or 8 items (120, 360, or 960 ms SOAs) or in the absence of a target. Each surprise stimulus was randomly chosen from the set of stimuli described above without replacement; thus, a given participant never experienced the same surprise stimulus twice. Each of the four surprise trial types was presented at a random point within each block with three restrictions. First, the initial surprise trial in each block was between trial 6 and 11. Second, there were at least three non-surprise trials between any two surprise trials. This restriction was intended to attenuate any immediate carryover effects from the post-surprise trial (Asplund, Todd, Snyder, & Marois, 2010) or inter-trial priming effects (Maljkovic & Nakayama, 1994). Third, the first four surprise trials of each session included one of each type, with an order that was counterbalanced across participants. This consideration ensured that we could examine habituation effects for each surprise-target lag.

The main experiment consisted of 360 trials divided into 9 blocks of 40, enabling us to characterize a wide range of habituation rates. Between blocks, participants were encouraged to take a short break. Before completing the main experiment, participants completed at least 6 practice trials. No surprises were presented during the practice trials. Most participants completed the session within 60 mins.

Statistical approach. Data preparation, visualizations, and statistical analyses for this and subsequent experiments were implemented in RStudio version 1.0.153 (R Foundation for Statistical Computing) running R version 3.2.4. Prior to conducting formal analyses on each experiment, we identified individuals who had poor baseline performance or who had numerous trials with multiple responses. Inferring surprise-related deficits from such individuals' performance would be difficult, so their data were removed from each sample. Each participant's detection performance on non-surprise trials was represented as a d-prime score, with extreme hit or false alarm rates accommodated by setting these values to 0.1 or 0.9 for the z-score transformation (Stanislaw & Todorov, 1999). In Experiment 1, d-prime values less than approximately 0.5 corresponded to performance indistinguishable from chance responding; we used this threshold for subsequent experiments as well. Inspection of the relevant histograms revealed that only a few individuals had numerous trials with multiple responses, whereas most individuals had 0-3. We therefore used 10 trials with multiple responses as a cutoff threshold. These data removals applied to both detection and reaction time analyses. For the latter, we also removed individual trials with reaction times over 5 seconds to minimize the effect of outliers.

For Experiment 1 and most subsequent experiments, we constructed and fit linear mixed-effects models (LMEMs). These models allowed us to examine trends in performance across time. Correct detection (hit) rates were assessed using logistic regression implemented with glmer() in the lme4 package (version 1.1.13) (Bates, Mächler, Bolker, & Walker, 2015). False alarms were assessed in the same way in a separate model. Reaction times were assessed using linear regression implemented with lmer(). For the models that assessed hits and reaction times, the factors were Trial Type (4 levels: Lag 1, 3, 8, or Probe only; categorical) and Time (block number; continuous). For subsequent experiments, the Time factor was set up by first labelling each trial with its serial position within the session; this assignment was done separately by Trial Type, after which the values were normalized to the range -0.5 to 0.5. The models for false alarms contained the same factors (Trial Type and Time), but the two Trial Type levels were Surprise only and No probe, no Surprise. Since all factors in each of these models were manipulated within-subjects, the random effects structure included by-subject random intercepts and by-subject random slopes for the Trial Type factor. Including by-subject random slopes for the Trial Type x Time interaction produced models that frequently failed to converge, so the random effects structure used is the maximum supported by the design and data (Barr, Levy, Scheepers, & Tily, 2013). For models comparing the results across experiments, we included each of the factors above, added a between-groups factor of Experiment, and used the same random effects structure as before. For all LMEMs, main effects and interactions were assessed using Type II Wald chisquare tests via the Anova() function in the car package (version 2.1.5) (Fox & Weisberg, 2011).

Post-hoc tests were conducted using the emmeans(), emtrends(), and contrast() functions in the emmeans package (version 1.3.4) (Lenth, Singmann, Love, Buerkner, & Herve, 2019). Pairwise differences between estimated marginal means and trends were assessed using t-distributions with the degrees of freedom based on Satterthwaite's method. To account for multiple comparisons, we applied Bonferroni correction to each set of post-hoc comparisons; each p-value was multiplied by the number of comparisons in the set.

Results

Twelve participants (24% of original sample) had either poor baseline detection performance (11), excessive trials with multiple responses (4), or both. After their removal, 37 participants remained in the final sample.

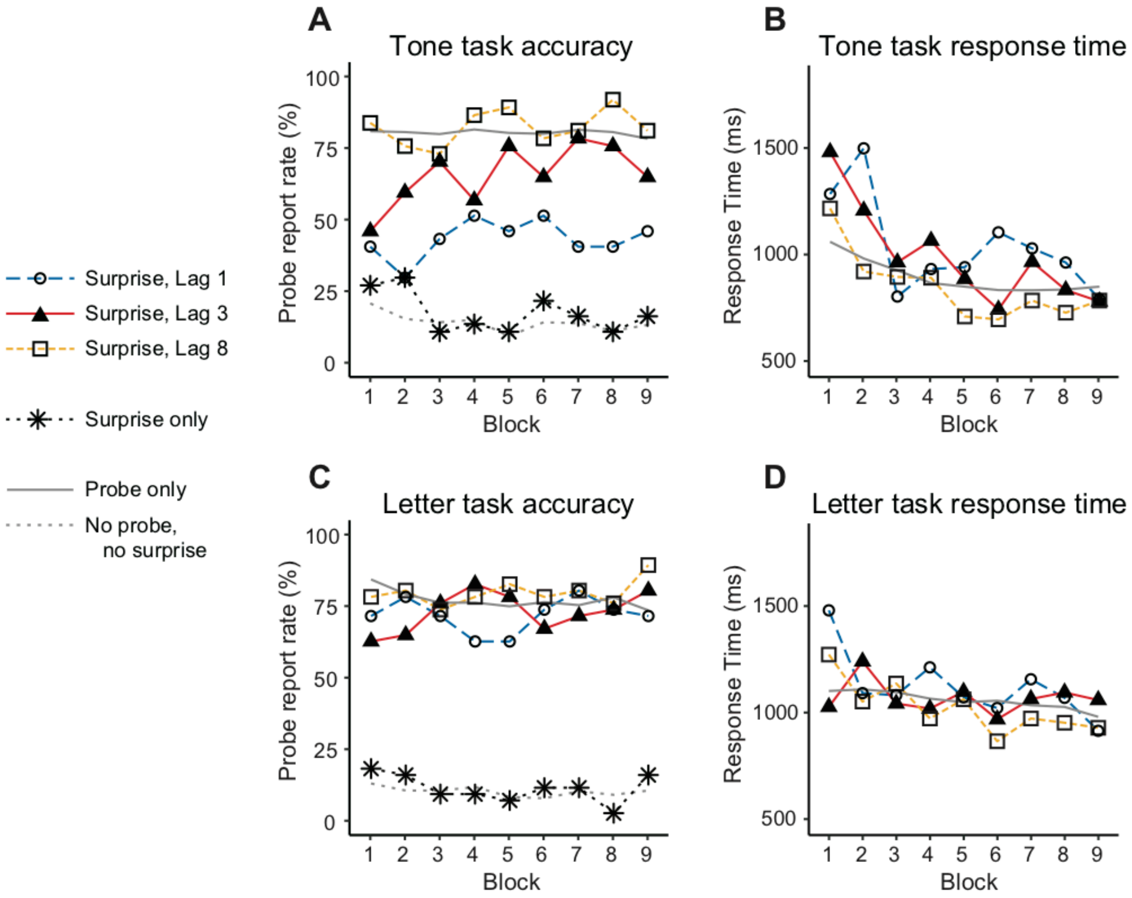


Figure 2. Average probe performance for each Trial Type by Block. (A) Probe report rate in Experiment 1A (tone task). Note that these values are false alarm rates in the "Surprise only" and "No probe, no surprise" categories. (B) Probe response time in Experiment 1A. (C) Probe report rate in Experiment 1B (letter task). (D) Probe response time in Experiment 1B. Error bars have been omitted for clarity.

Probe report. For visualization purposes, the average probe report rate was calculated for each Trial Type and block (Figure 2A). Linear Mixed-Effect Models (LMEMs, see above) with within-subject factors of Time (Block 1 to 9) and Trial Type (Probe only, Lag 1, Lag 3, Lag 8) were used for analysis. For hits, there was a significant main effect of Trial Type and a significant Time x Trial Type interaction (Table 1). The hit rate for Lag 1 was significantly lower than for other Trial Types (*p*s < .001). The Lag 3 rate was lower than the Lag 8 rate, though not significantly (*p* = .059). The other two pairwise differences were not significant (*p*s > .10). The Lag 3 deficit habituated more quickly (higher Time slope) than Probe only (*p* = .014), which was due to a significant positive slope for Lag 3 (linear change across experiment: mean = 15.9%, SE = 5.5%; *p* = .015) and no significant change over Time for Probe only (*p* = 1). No other slopes or slope differences were significant (*p*s > .85). To better understand the habituation at Lag 3, we compared Lag 3 and Probe only hit rates for each block with Wilcoxon signed ranks tests owing to the binary dependent measure. After Bonferroni correction, only the deficit in block 1 (mean = 34.9%, SD = 50.7%) remained significant (*p* = .013; *p*s > .25 for all others).

For false alarms, there was a significant main effect of Time ($\chi^2(1) = 13.00$, *p* < .001), as the false alarm rate decreased across blocks. Neither the main effect of Trial Type nor the Time x Trial Type interaction was significant (*p*s > .46). Surprise only trials, particularly in early blocks, contained numerically more false alarms, so it is unlikely that either Surprise effects or the habituation of Lag 3 effects were due to response bias. By the same token, there were relatively few responses to the Surprise stimulus itself.

Table 1. Statistical results for LMEMs assessing probe hit rates in Experiments 1A and 1B as a function of Time and Trial Type. The LMEM comparing the two experiments included Experiment as a factor as well. Significant effects are indicated in **bold**. dfs = degrees of freedom.

	Expt. 1A (tone task)	Expt. 1B (letter task)	Both Expts.

	dfs	χ^2(dfs)	*p*	χ^2(dfs)	*p*	χ^2(dfs)	*p*
Time	1	<.001	.99	**21.24**	**<.001**	**12.31**	**<.001**
Trial Type	3	**55.65**	**<.001**	7.69	.053	**55.23**	**<.001**
Experiment	1					0.01	.92
Time × Trial Type	3	**11.33**	**.010**	**11.70**	**.008**	**21.99**	**<.001**
Trial Type × Experiment	3					**35.96**	**<.001**
Time × Experiment	1					**8.72**	**.003**
Time × Trial Type × Experiment	3					0.41	.94

In Surprise-induced Blindness, the Lag 3 deficit's habituation was often complete within 2-4 trials (Asplund, Todd, Snyder, Gilbert, et al., 2010; Asplund, Todd, Snyder, & Marois, 2010). To test for such rapid effects in the present experiment, we examined performance during the Lag 3 trials in the first block. Each participant experienced one Lag 3 trial in the first four Surprise trials. Therefore, we calculated the detection rate as a function of the ordinal number of the Lag 3 trial relative to other Surprise trials (1st: 29%, 2nd: 56%, 3rd: 27%, 4th: 60%). No clear trend was evident in the detection rates, and a non-parametric Cochran Q test found no significant differences across the four categories (Q(3) = 3.46, $p = .33$). Surprise-induced Deafness' habituation appears to be gradual, such that variance overwhelms within-block habituation trends.

Reaction times. Only trials with correct Probe detection were used for reaction time (RT) analyses. For visualization purposes, mean RTs for hits were calculated for each Trial

Type and block (Figure 2B). Formal analysis revealed significant main effects of Time and Trial Type as well as a significant Time x Trial Type interaction (Table 2). RTs in the Lag 1 and Lag 3 conditions were significantly longer than in the other two conditions (*p*s < .001), whereas other pairwise comparisons were not significant (*p*s > .91). Although RTs significantly decreased for all Trial Types across Time (*p*s < .001), save Lag 1 (*p* = .17), only the Lag 3 and Probe only trials had significantly different slopes (*p* = .001; *p*s > .28 for all others).

Table 2. Statistical results for LMEMs assessing reaction times for probe hits in Experiments 1A and 1B as a function of Time and Trial Type. The LMEM comparing the two experiments included Experiment as a factor. Significant effects are indicated in bold. dfs = degrees of freedom.

		Exp. 1A (tone task)		Exp. 1B (letter task)		Both Expts.	
	dfs	χ^2(dfs)	*p*	χ^2(dfs)	*p*	χ^2(dfs)	*p*
Time	1	**174.84**	**<.001**	**52.51**	**<.001**	**199.88**	**<.001**
Trial Type	3	**34.20**	**<.001**	5.22	.16	**30.32**	**<.001**
Experiment	1					**5.12**	**.024**
Time × Trial Type	3	**16.18**	**.001**	**9.89**	**.020**	**13.04**	**.005**
Trial Type × Experiment	3					**14.95**	**.002**
Time × Experiment	1					**14.95**	**.002**

Time × Trial Type × Experiment	3	**12.61**	**.006**

Experiment 1B: SiD in a letter detection task

Method

Experiment 1B was conducted concurrent with Experiment 1A, and the experiments were identical with the following exceptions. Fifty-three participants (21 males, 3 gender not reported; mean age = 21.78 years, age range = 19-30) searched for a spoken letter "X" probe amongst spoken letter distractors. Surprise stimuli were pure tones, and each participant experienced a given tone frequency only once.

Results

Eight participants (15% of the sample) had either poor baseline detection performance (4), excessive trials with multiple responses (5), or both. After their removal, 45 participants remained in the final sample.

Probe report. The pattern of results bore some similarities to those of Experiment 1A, though the effects were generally weaker (Figure 2C). The main effect of Time and the Trial Type x Time interaction were significant, though the main effect of Trial Type did not reach significance (Table 1). Unlike in the previous experiment, hit rates did not significantly differ across Trial Types (ps > .16, except $p = .053$ for the difference between Lag 1 and Lag 8). Lag 3 and Probe only slopes across Time did not significantly differ ($p = .063$), with the trend driven by a numerically small but significant decrease in the Probe only hit rate ($p < .001$). No other slopes or slope differences were significant (ps > .29). The hit rates for Lag 3 and Probe only trials also did not significantly differ within any blocks (ps > .21).

For false alarms, there was a significant main effect of Time ($\chi^2(1) = 4.22$, $p = .040$), with the rate decreasing across blocks. There was no significant main effect of Trial Type and no significant Time x Trial Type interaction (*p*s > .51).

As we had for Experiment 1A, we analyzed the probe detection rate for the first block's Lag 3 trials as a function of Surprise stimulus experience. Again, there was no apparent trend to performance (1st: 64%, 2nd: 73%, 3rd: 75%, 4th: 36%), and a Cochran Q test found no significant difference across these categories ($Q(3) = 4.49$, $p = .21$).

Reaction times. The pattern of results was broadly similar to that in Experiment 1A, albeit with weaker Surprise effects (Figure 2D). The main effect of Time and the Trial Type x Time interaction were both significant, whereas the main effect of Trial Type was not (Table 2). Participants became faster overall as the experiment progressed, and there were significant RT decreases for all Trial Types (*p*s < .004) save Lag 3 ($p = 1$). None of the pairwise comparisons of slopes were significant (*p*s > .15).

Probe report across Experiments 1A and 1B. Performance across the experiments was assessed by adding a factor of Experiment to our LMEMs (Table 1). Both the Experiment x Trial Type and Experiment x Time interactions were significant, whereas the three-way Time x Trial Type x Experiment interaction was not. Performance for Lag 1 was lower in Expt. 1A than in 1B ($p < .001$; Figures 2A, 2C), but no other Trial Type differences were significant (*p*s > .13). Probe only trials differed in their slopes across experiments ($p = .005$), driven by the aforementioned decrease in performance across Time in Experiment 1B. No other Trial Type slopes significantly differed across experiments (*p*s > .29). There were more false alarms in the tone task (Experiment 1A), though this difference did not produce a significant main effect of Experiment ($\chi^2(1) = 2.85$, $p = .092$). For false alarms, no interactions with Experiment were significant (*p*s > .30).

Reaction times across Experiments 1A and 1B. All main effects and interactions were significant (Table 2), with the main effect of Experiment due to longer RTs in Expt. 1B.

RTs were marginally longer for Lag 1 in Expt. 1B (*p* = .056), though there were no significant differences for other Trial Types (*p*s > .11). RT slopes were significantly different across Experiments for Probe only and Lag 3 (*p*s < .001; *p*s > .54 for others). Whereas RTs decreased for Probe only trials in both experiments, Lag 3 RTs decreased only for Expt. 1A (see experiment-specific post-hoc tests above).

Discussion

In this first experiment, we demonstrated that relatively rare and unexpected auditory stimuli can induce a detection deficit. This deficit, which we term Surprise-induced Deafness (SiD), has many characteristics that are similar to its visual counterpart, Surprise-induced Blindness (SiB; Asplund, Todd, Snyder, Gilbert, et al., 2010). In both SiD and SiB, the impairment in probe detection depends on the time between the Surprise stimulus and the probe, with larger deficits at Lags 1 and 3. The effects in these two conditions can be distinguished by the habituation rates, with more rapid reduction of the Lag 3 effects in each paradigm. Nevertheless, we also found substantial differences across the two modalities. SiD was a less severe deficit initially, but the rate of habituation was also considerably slower than SiB's. Furthermore, SiD only partially generalized to different classes of stimuli. Below, we consider each of these features for the Lag 1 and Lag 3 deficits, as well as how they will be explored in subsequent experiments in this study.

For the tone task (Experiment 1A), there was a persistent and severe deficit at Lag 1. One possibility is that this deficit reflects perceptual processes, such as auditory masking (Brosch & Schreiner, 1997; Jesteadt, Bacon, & Lehman, 1982; Moore & Glasberg, 1981; Pastore, Harris, & Goldstein, 1980). Auditory masking persists across hundreds or even thousands of trials (Delahaye, Fantini, & Meddis, 1999). Furthermore, a dynamic letter may be a more effective mask for a tone than vice versa, which would be consistent with the lack of Lag 1 effects in Experiment 1B. (It must be noted, however, the effects at Lag 3 did not reach statistical significance either.) Alternatively, the Lag 1 deficit in Experiment 1A may represent an attentional effect that habituates slowly, akin to what was observed in Surprise-induced Blindness (Asplund, Todd, Snyder, Gilbert, et al., 2010). We return to this possibility in the General Discussion.

The Lag 3 deficit, by contrast, was evidenced in both versions of the experiment. It also habituated with repeated exposures, most clearly in Experiment 1A. Such habituation is consistent with attentional reallocation due to an orienting response (Kahneman, 1973; Pavlov, 1927; Sokolov, 2002). As Lag 3 corresponds to an SOA of 400 ms, the deficit's timecourse within each trial is also consistent with delayed attentional capture following a surprising event (Asplund, Todd, Snyder, Gilbert, et al., 2010; Horstmann, 2005, 2015; Theeuwes, Atchley, & Kramer, 2000). In addition, it is consistent with the maximal attentional blink effects that occur between 200-500 ms after the inducing target item. This timing is characteristic of the visual AB (Dux & Marois, 2009), and it is frequently found for the auditory AB as well (Arnell & Jolicoeur, 1999; Martens et al., 2009; Tremblay et al., 2005). Taken together, the features of the Lag 3 deficit suggest that it is sensitive to informational surprise, unlike the Lag 1 effects. Surprise-induced Deafness (SiD) therefore refers to the former effect, and we investigate Lag 3 effects in subsequent experiments.

Despite the robust habituation of SiD, the rate of habituation was slower than we had anticipated. Based on Surprise-induced Blindness (Asplund, Todd, Snyder, Gilbert, et al., 2010) and auditory novelty-P3 habituation rates (Friedman et al., 2001), we had expected complete attenuation after 2-6 presentations of the surprising stimulus. Attenuation in Experiment 1A, however, was not convincingly complete until at least 10-15 stimuli had been experienced. One possibility is that the Surprise stimuli were perceived as heterogenous, and such varying stimuli yield slower and incomplete habituation (Asplund, Todd, Snyder, Gilbert, et al., 2010; Asplund, Todd, Snyder, & Marois, 2010). We explore the effects of same and varying Surprise stimuli in Experiments 2 and 3. These experiments also allow us to explore the possibility that participants anticipate and inhibit the Surprise stimuli so that they have reduced effects over repeated presentations.

Finally, the differences between Experiments 1A and 1B suggest that stimulus features contribute to SiD. One possibility is that letters are more salient because they more meaningful than pure tones. We investigate the role of the Surprise stimuli's content in Experiment 4. Alternatively, letters may have different effects in RAP streams because they are dynamic—a series of tones is arguably more like a series of visual stimuli, each of which is static—or because our stimulus equalization method did not fully work as intended. We use a different method in Experiment 5.

Experiment 2: A simplified SiD paradigm

Having demonstrated the Surprise-induced Deafness effect, we next developed a simplified and abbreviated version of the paradigm. We focused on the Lag 3 effects, which evidenced both a strong initial deficit and habituation across Surprise trials. We used this new paradigm for the two aims of Experiment 2. First, we sought to replicate the salient features of SiD in a paradigm that could be adapted for use on Amazon Mechanical Turk (AMT), an online platform that can be used to conduct cognitive psychology experiments (Crump, McDonnell, & Gureckis, 2013). Given that SiD is calculated from rather few trials, a large sample size would be needed for subsequent investigations into higher-order effects. Second, we informally tested whether SiD would more quickly habituate when the same Surprise stimulus was used for each Surprise trial. Successive presentations of the Surprise stimulus would contain less novelty, potentially reducing its salience. Surprise-induced Blindness has been found to habituate more rapidly under such circumstances (Asplund, Todd, Snyder, Gilbert, et al., 2010; Asplund, Todd, Snyder, & Marois, 2010). In this experiment, we merely tested whether significant deficit habituation, perhaps even its complete elimination, could be observed in a simplified and abbreviated paradigm. In Experiment 3, we explored habituation rates formally using AMT.

Method

The procedure for Experiment 2 was broadly similar to Experiment 1's, with numerous specific changes to address the new experiment's goals and to improve the SiD paradigm. Twenty-two participants (7 males, mean age = 20.55, age range = 18-24) completed a 20-minute session of 120 trials in three blocks. The Lag 1 and 8 trials from Experiment 1A became Lag 3 trials in the present experiment, thereby producing the same number of Lag 3 Surprise trials as in Experiment 1 (9). Given the medium-to-large effect size for the first block's Lag 3 deficit in Experiment 1A (Cohen's d=0.69), a sample size of 20 would yield 80% power (albeit for parametric statistical results). We also anticipated significant, if not complete, habituation within the session. Each participant was presented with the same Surprise stimulus in each Surprise trial. The Surprise stimuli were spoken letters, with each assigned randomly to a participant (e.g., 'A' for the first participant, 'B' for the second participant etc.).

Instead of equilibrating stimulus volumes by matching maximum amplitudes, stimulus equilibration was achieved by matching their summed absolute amplitude values (sound envelopes). Before each trial, participants could play the probe as many times as they wished. Due to the multiple responses in some trials in Experiment 1, participants were explicitly warned not to respond more than once per trial if multiple responses were detected.

Results and Discussion

One participant (5% of the sample) was removed from further analysis owing to poor baseline detection performance, leaving 21 participants in the final sample. Warning messages were effective at reducing the incidences of multiple responses; no participant had more than four trials with multiple responses.

For visualization purposes, the average probe report rate was calculated for each Trial Type and block (Figure 3A). The Time x Trial Type interaction for hits was significant ($\chi^2(1) = 10.63$, $p = .001$), though neither main effect was (*p*s > .73). Performance increased across Time for Lag 3 trials ($p = .004$) but not for Probe only trials ($p = .67$). When assessed with Wilcoxon signed ranks tests, the hit rate for Lag 3 was significantly lower than for Probe only in Block 1 ($p < .049$), but subsequent blocks did not differ (*p*s > .35). Within the first block, the hit rate by Surprise trial (1st: 55%, 2nd: 59%, 3rd: 69%, 4th: 68%) did not significantly differ (Cochran Q(3) = 0.93, $p = .82$), consistent with the gradual habituation of a modest initial deficit. For false alarms, there was a significant main effect of Time ($\chi^2(1) = 6.89$, $p < .009$), as the false alarm rates decreased across blocks. Neither the main effect of Trial Type nor the Time x Trial Type interaction was significant (*p*s > .20). As such, the habituation of the Lag 3 Surprise effects was not due to response bias or participants mistakenly reporting the Surprise stimulus as the probe.

For reaction times, there was a significant main effect of Trial Type ($\chi^2(1) = 5.21$, $p = .022$), with Surprise stimuli causing slower responses. There was also a marginally significant main effect of time ($\chi^2(1) = 3.15$, $p = .076$), with responses becoming faster over the course of the experiment. The Time x Trial Type interaction was not significant ($\chi^2(1) = 0.75$, $p = .39$). Although the reaction time effects largely aligned with the detection results, we focused on the latter for subsequent experiments because they are the critical measure for related effects such as Surprise-induced Blindness (Asplund, Todd, Snyder, Gilbert, et al., 2010) and the auditory Attentional Blink (Horváth & Burgyán, 2011; Mondor, 1998; Shen & Mondor, 2006; Vachon & Tremblay, 2005, 2006). More generally, Experiment 2 showed that a simplified and abbreviated SiD paradigm still contained the effect’s key features, thereby providing us a path towards addressing additional questions using Amazon Mechanical Turk (AMT).

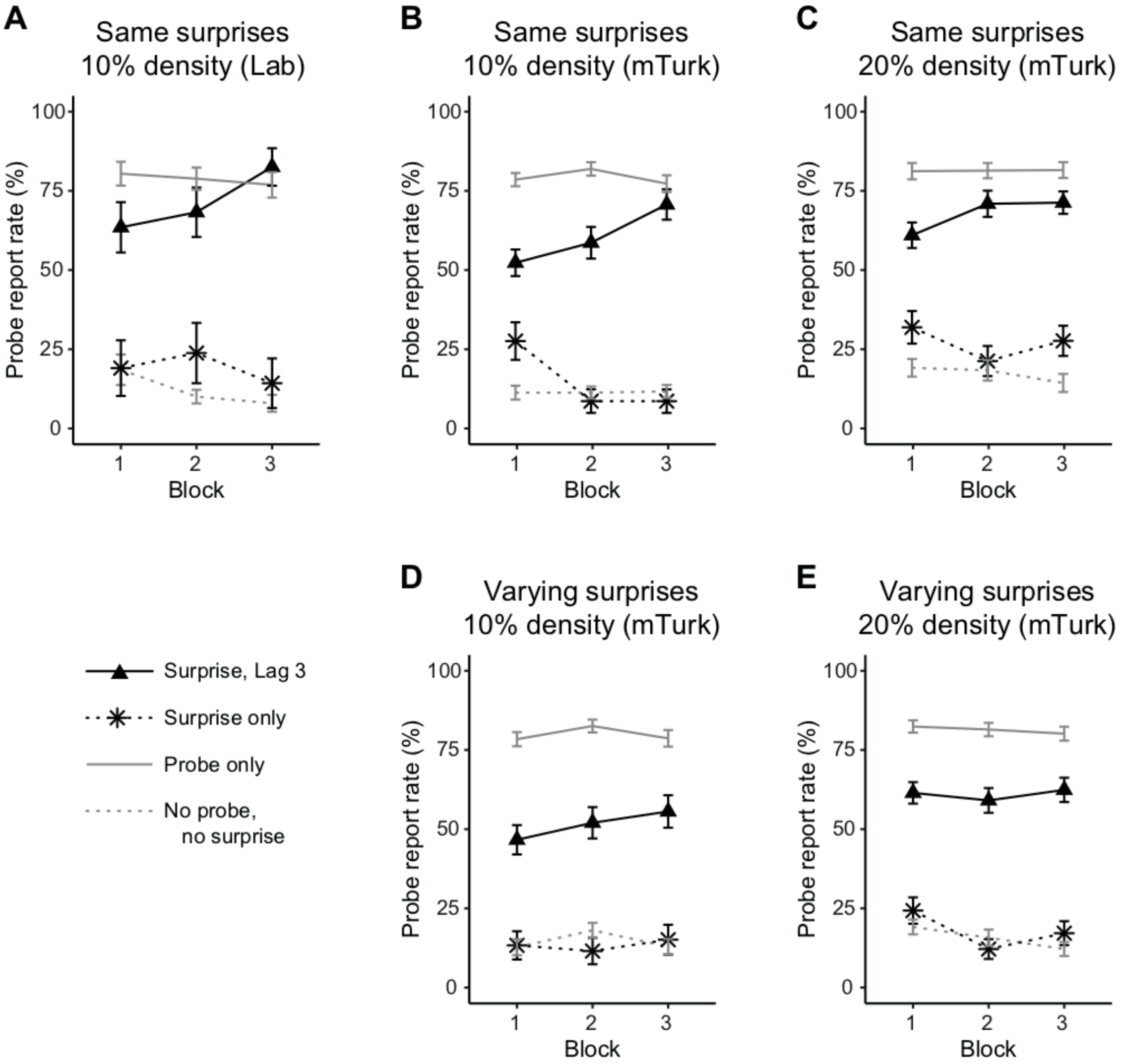


Figure 3. Mean probe report rates for each Trial Type plotted by block in Experiment 2 (Panel A) and Experiment 3 (Panels B-E). Note that these values are false alarm rates in the “Surprise only” and “No probe, no surprise” categories. Panels A and B show the effect of the testing platform (lab-based or online), whereas other comparisons illustrate the effects of Surprise stimulus homogeneity (Panels B and C versus D and E) and density (Panels B and D versus C and E). For all panels, error bars represent ±1 standard error of the mean (SEM).

Experiment 3: Surprise stimulus novelty and frequency

In Experiment 2, we used a simplified paradigm to replicate the key features of SiD, both the overall deficit and its habituation. In the present experiment, we leveraged this simplified design to better understand the effects of Surprise stimulus novelty and frequency of appearance on SiD magnitude using Amazon Mechanical Turk (AMT). Both novelty and frequency are factors that affect stimulus salience (Downar et al., 2002), but one or both factors may be independent of the mechanisms that induce SiD.

The violation of expectations ("surprise") is essential for auditory distraction to occur in many contexts (Parmentier, Elsley, Andrés, & Barceló, 2011; Vachon, Hughes, & Jones, 2012). The information content of task-irrelevant stimuli affects whether and how much attention they will grab. Expectations are built up dynamically as participants learn about the task and events that are likely to occur within it. If Surprise stimuli occur frequently, each violates expectations less and carries less information. Accordingly, an increase in the density of Surprise stimuli relative to other items can reduce or abolish surprise effects (Asplund, Todd, Snyder, Gilbert, et al., 2010; Meyer et al., 1991; Niepel et al., 1994). Similarly, when Surprise stimuli do not change from trial to trial, more specific expectations can be developed. As such, Surprise effects may habituate more rapidly (see Experiment 2). In contrast, heterogenous Surprise stimuli may lead to effects that incompletely habituate (Asplund, Todd, Snyder, Gilbert, et al., 2010; Asplund, Todd, Snyder, & Marois, 2010).

Note that these effects of density and heterogeneity manipulations are not orthogonal, and either manipulation can affect both the overall magnitude of Surprise-induced Deafness and its habituation rate or completeness. Indeed, one would expect that the first instance of SiD would be insensitive to these manipulations, as the item would be new and rare, with performance developing from this common point. Accordingly, we had a similar hypothesis for each manipulation: If Surprise-induced Deafness is caused by the surprising aspect of the inducing stimuli, the deficit will be sensitive to changes in participant expectation. We anticipated the largest deficits and slowest habituation rates when the Surprises varied across trials and were most infrequent. In addition to testing this hypothesis, we compared the results from Experiments 2 and 3 to understand the similarities and differences across experimental platforms.

Method

Participants. Three hundred and five participants (see Table 3) with normal hearing participated for payment ($1) via Amazon Mechanical Turk (AMT), an online crowdsourcing platform. On AMT, experimenters ("requesters") post human intelligence tasks (HITs) that anonymous potential participants can choose to complete. To ensure the quality of data, participants were required to have 90% of their previous HITs approved by the requester and to have at least 1,000 HITs completed (Grysman, 2015; Peer, Vosgerau, & Acquisti, 2014; Shapiro, Chandler, & Mueller, 2013; Summerville & Chartier, 2013).

The use of AMT made quickly collecting a large sample size feasible. As we had few expectations about the effect sizes for our between-group comparisons for heterogeneity and density, we planned to collect data until each group had at least 40 participants (double the total number of participants in Experiment 2); recruitment was faster than anticipated, so we collected a rather larger sample. Each participant was randomly allocated to one of four groups (Table 3). Participants within each group experienced different Surprise stimulus properties, two levels each for Heterogeneity and Density. For Heterogeneity, participants could experience either the same Surprise stimulus for each Surprise trial (Same) or a different stimulus for each trial (Varying). For Density, Surprise trials were either 10% (4 of 40) or 20% (8 of 40) of the total number of trials within each of three blocks.

Table 3. Participant information for Experiment 3. The data from participants with low probe

detection performance (d’<0.5) were removed from the sample. Group sizes are unequal because each participant was assigned a group randomly. NR = not reported.

Heterogeneity	Density	Recruited sample size	Ages	Removed	Final sample size
Same	10%	70 (36 males)	30.93 (21-62), NR=1	12 (17%)	58
Same	20%	64 (36 males)	30.31 (21-60), NR=2	17 (27%)	47
Varying	10%	79 (43 males)	31.78 (20-52), NR=1	23 (29%)	56
Varying	20%	92 (58 males)	30.85 (18-65), NR=1	22 (24%)	70
Total		305 (173 males)	30.97 (18-65), NR=5	74 (24%)	231

Stimuli and Apparatus. The experiment itself was conducted on Amazon Mechanical Turk (AMT). Previous studies have shown that AMT-based experiments can replicate the visual attentional blink and similar cognitive tasks (Crump et al., 2013). Once individuals on AMT had agreed to participate in our experiment, they were redirected to an online survey system (Qualtrics, 2005) through which stimuli were presented and the task flow was controlled. At the end of the experiment, participants received instructions to enter a unique code generated by Qualtrics in the AMT HIT to verify that they completed the study to receive payment. Qualtrics restrictions were set to allow one response per AMT worker ID to provide protection against participants completing the study multiple times.

The presented stimuli were identical to those used in Experiment 2, except that we expanded the set of Surprise stimuli. In addition to the sounds used before, the Surprise stimulus set included the following sounds: alarm, balloon popping, car horn, cat meowing, cough, cowbell, dog barking, giggle, hiccup, hi-hat cymbal, lightbulb breaking, mosquito, plunger, slide whistle, slurping, snare drum, sneeze, tongue popping, “one”, and “two”. The spoken digits were recorded from a female native English speaker, whereas the other sounds were drawn from freesound.org (Font, Roma, & Serra, 2013). All sounds were compressed to 110 ms each, with care taken to preserve intelligibility.

Task procedures. The task procedures were identical to Experiment 2 with the following exceptions. Foremost, each participant was allocated to one of four groups with different Surprise trial features, as described above. All responses were unspeeded, with the probe response made at the conclusion of the stream. Before beginning the main experimental blocks, participants completed six practice trials of the task at a slightly slower speed (120 ms SOAs) and another 6 practice trials at normal speed (110 ms SOAs). These practice trials contained feedback. During the practice period, participants were also instructed to adjust the volume to a comfortable level, and headphones were recommended.

Owing to randomization limitations with Qualtrics, a limited number of trial orders were used. All participants in the Same 10% and Varying 10% groups experienced the same sequence of trials (only the Surprise stimuli were different), whereas the Varying 20% and Same 20% groups used two different orders. All trial sequences were visually inspected to ensure a reasonable distribution of trial types across time. Critically, the number of trials for each Trial Type was consistent across blocks. They also followed the same restrictions used in Experiment 1 and 2, with the exception that at least two (instead of three) trials without a Surprise stimulus had to occur between any two Surprise trials.

Results and Discussion

Similar to the previous experiments, participants whose baseline probe detection rate was low (d' < 0.5) were excluded from further analyses (Table 3). We investigated effects within each group first, and then compared across groups to understand the effects of Surprise stimulus Heterogeneity and Density.

Probe reports within each group. For visualization purposes, the average probe report rate for each of the four groups was calculated for each Trial Type and block (Figure 3B-E). For hits, the main effect of Trial Type was significant in each group (Table 4), demonstrating that Surprise stimuli impaired performance. Three of the four groups also had significant Trial Type x Time interactions (Table 4). The groups with unchanging Surprise stimuli evidenced clear effect habituation across repeated Surprise stimulus presentations: Hit rates significantly increased over Time for Surprise trials (Same 10% group: p = .001, Same 20% group: p = .005), whereas the hit rate for Probe only trials either decreased (p = .020) or remained unchanged (p = .55), respectively. Habituation effects were less clear for the Varying groups. The Varying 10% group showed a marginal increase across Time for Surprise trials (p = .079) and no change for Probe only trials (p = .29). The Trial Type x Time interaction was marginal for the Varying 20% group, but the main effect of Time was significant. This pattern was due to a significant decrease in the hit rate during Probe only trials (p = .007) and no significant change for Surprise trials (p = 1.00).

Table 4. Statistical results for LMEMs assessing probe hit rates in Experiment 3 as a function of Time and Trial Type. Significant effects are indicated in **bold**.

	Same 10%		Same 20%		Varying 10%		Varying 20%	
	$\chi^2(1)$	*p*	$\chi^2(1)$	*p*	$\chi^2(1)$	*p*	$\chi^2(1)$	*p*
Trial Type	**38.58**	**<.001**	**23.62**	**<.001**	**63.49**	**<.001**	**103.67**	**<.001**

Time	1.04	.31	0.37	.54	0.32	.57	**5.01**	**.025**
Trial Type x Time	**19.07**	**<.001**	**9.31**	**.002**	**6.06**	**.014**	3.65	.056

Probe detection performance for each Trial Type began at a similar level within each group (Figure 3). Indeed, accuracy for the first Surprise trial (which always contained a probe item) was similar across the four groups (Same 10%: 48%; Same 20%: 47%; Varying 10%: 57%; Varying 20%: 60%). These values did not significantly differ (Cochran Q(3) = 2.97, *p* = .40).

We also examined the false alarm rate within each group. As we had found in Experiments 1 and 2, this rate either decreased over Time (Same 20%: $\chi^2(1) = 6.86$, $p = 009$; Varying 20%: $\chi^2(1) = 20.90$, $p < .001$) or did not significantly change. A significant Trial Type x Time interaction for Same 10% ($\chi^2(1) = 7.73$, $p = .005$) likely reflects a similar effect, as the rate significantly decreased only for Surprise trials ($p = .004$; $p = 1.00$ otherwise). The higher false alarm rate for Surprise trials in the Same 20% group ($\chi^2(1) = 13.89$, $p < .001$) was both numerically small and apparently an isolated oddity. No other effects were significant ($\chi^2(1)$s < 2.14, *p*s > .14). Overall, the false alarm rate was similar to our lab-based results. The results suggest that our Surprise effects were not due to response bias or participants mistakenly reporting the Surprise stimulus as the probe, even online using AMT.

Formal comparisons across groups. Our analyses showed robust Surprise-induced Deafness effects within each group, though the results also suggested differences across them. To test these observations statistically, we constructed and fit a LMEM with Heterogeneity (homogeneous vs. heterogeneous) and Density (10% vs. 20%) as between-subject variables, and Time (continuous) and Trial Type (Probe only, Lag 3 Surprise) as within-subject variables.

Table 5. Statistical results for LMEMs assessing probe hit rates in Experiments 1A and 1B as a function of Time and Trial Type. The LMEM comparing the two experiments included Experiment as a factor as well. Significant effects are indicated in **bold**. dfs = degrees of freedom.

	$\chi^2(1)$	*p*
Time	2.62	.11
Trial Type	**151.17**	**<.001**
Heterogeneity	1.40	.24
Density	**38.07**	**<.001**
Trial Type × Time	**27.59**	**<.001**
Heterogeneity × Time	1.21	.27
Heterogeneity × Trial Type	**17.67**	**<.001**
Density × Time	0.68	.41
Density × Trial Type	**8.85**	**.003**
Heterogeneity × Density	2.22	.14
Heterogeneity × Trial Type × Time	2.84	.092
Density × Trial Type × Time	1.35	.24
Heterogeneity × Density × Time	1.45	.23

Heterogeneity × Density × Trial Type	0.15	.70
Heterogeneity × Density × Trial Type × Time	0.07	.78

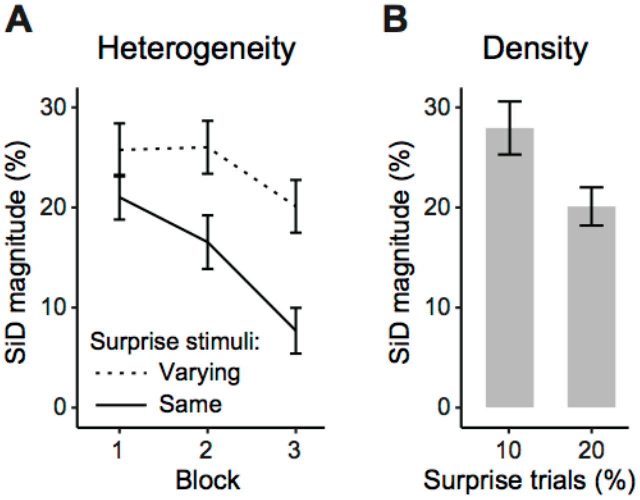


Figure 4. Summary of effects from the Heterogeneity and Density manipulations. Panel A: Mean SiD magnitude for homogeneous and heterogeneous condition as a function of Time in Experiment 3. Solid and dotted lines represent the SiD magnitude of homogeneous and heterogeneous conditions respectively. Panel B: Mean SiD magnitude as a function of Density in Experiment 3. To equate the number of Surprise stimuli contributing to each bar, the data are from the first two blocks of the 10% group and the first block of the 20% group, yielding six contributing Surprise trials per participant for each. For both panels, error bars represent ±1 standard error of the mean (SEM).

Surprise stimulus heterogeneity. For visualization purposes, “SiD magnitude” was calculated by subtracting the Surprise trials’ hit rate from the Probe only trials’ hit rate (Figure 4A). The LMEM analysis revealed a significant interaction of Heterogeneity x Trial Type, indicative of greater SiD when Surprise stimuli varied across trials (Table 5). Follow-up analyses showed that Surprise trials were affected by the Heterogeneity manipulation ($p = .030$), whereas Probe only trials were not ($p = .92$). There was also a marginally significant three-way interaction of Heterogeneity x Trial Type x Time (Table 5). The Trial Type x Time interaction was significant for both the Same ($p < .001$) and Varying ($p = .003$) groups. The habituation effect was more convincing for the former, however, as Surprise trial performance improved over Time for the Same group ($p < .001$) but only marginally increased for the Varying group ($p = .079$; see also Figure 3B and 3C versus 3D and 3E). Performance for Probe only trials significantly decreased across Time in both groups (*p*s $< .018$).

Overall, changing Surprise stimuli led to larger deficits, with a trend towards slower effect habituation as well. These effects were found despite the Varying group starting with numerically smaller effects in the first Surprise trial (see above). These results are consistent with Surprises being distracting because of the information they contain. Even as individuals learn that the task involves unexpected events in general, their novel identities cause a reallocation of attention to the detriment of probe detection. Although this reallocation contains stimulus-driven components, it may also involve goal-directed or other forms of attentional control, an idea to which we return in the General Discussion.

Surprise stimulus density. For visualization purposes, the effect of Density was investigated by comparing SiD magnitude across the 10% and 20% groups (Figure 4B). The LMEM analysis revealed a significant interaction of Density x Trial Type (Table 5). Although there was also a main effect of Density, the difference in hit rates as a function of Density was more pronounced for Surprise trials (10% group $M = 0.56$ vs. 20% group $M = 0.63$; $p < .001$) than for Probe only trials (10% group $M = 0.79$ vs. 20% group $M = 0.81$; $p < .001$). We conclude that more frequent Surprise trials led to reduced SiD. Since there were no higher-order interactions with Density (Table 5), we found no evidence of habituation differences.

This null result could represent a model specification issue, as Time did not account for the different number of Surprise trials across Density conditions. When running a revised model based on a matched number of Surprise trials, however, there were still no significant interactions that involved both Time and Density. We therefore conclude that our Density manipulation primarily affected the magnitude of SiD, again consistent with the notion that attention is reallocated towards stimuli that carry more information. Note that the "violation of expectation" behind such information is relative; participants undoubtedly came to expect Surprise stimuli in both the 10% and 20% Density groups, but they were relatively rarer and therefore less expected in the former.

Comparison of lab-based and online results. The lab-based (Experiment 2) and online (Experiment 3, Same 10% group) experimental results were compared in a single LMEM by adding a between-groups factor of Experiment. Neither the main effect of Experiment ($\chi^2(1) = 0.12, p = .73$), the Time x Experiment interaction ($\chi^2(1) = 0.52, p = .47$), nor the Time x Trial Type x Experiment interaction ($\chi^2(1) = 0.72, p = .40$) was significant. The Trial Type x Experiment interaction ($\chi^2(1) = 6.21, p = .013$), however, was significant. Performance for Surprise trials was marginally lower in Experiment 3 ($p = .083$), but the performance for Probe only trials did not significantly differ across Experiment ($p = .56$). Therefore, the interaction reflects larger overall SiD effects in the AMT version (Experiment 3), but no difference in baseline Probe hit rates (see Figures 3A and 3B). This difference could be due to the less controlled environment for Experiment 3 participants, perhaps because of headphone differences. It is worth noting, however, that the SiD magnitude in Experiment 3 (Same 10% group) is rather similar to that observed in Experiment 1A (Figure 2). For false alarms, only the three-way interaction was significant ($\chi^2(1) = 7.44, p = .006$; other effects: $\chi^2(1) < 1.80, p > .18$). The Time x Experiment interaction for Surprise trials was marginally significant ($p = .083$), whereas it was not for trials with neither a probe or Surprise stimulus (*p* = .56). Taken together, these results suggest that SiD effects are largely comparable across lab-based and online experiments.

Experiment 4: Behavioral relevance of Surprise stimuli

For our tone tasks in the first three experiments, each Surprise stimulus was a spoken letter (Experiments 1-3) or an environmental sound (Experiment 3). Stimuli in these categories may capture attention more easily than other types, as they are broadly relevant to behavior (Egeth & Yantis, 1997). Indeed, it is possible that the weaker effects observed with the letter task (Experiment 1B) were in part due to the tone Surprise stimuli. In this experiment, we tested whether different categories of Surprise stimuli, constructed sine-wave tones or meaningful auditory objects, would capture attention primarily because of their deviance from expectation, thereby yielding comparable SiD effects. Conversely, if broad behavioral relevance of the Surprise stimuli is an important factor in inducing SiD, we might expect reduced or even absent effects with less meaningful tones. A secondary aim was to test whether SiD would still be observed when baseline detection performance was high for virtually all participants.

Method

The current experiment was identical to the Heterogenous 20% condition of Experiment 3 with the following exceptions. Thirty-nine participants (25 males, mean age = 28.74 years, age range = 22-42) completed the task; we aimed for a sample size twice that of Experiment 2, but had a data recording error for one individual. The experimental task included 12 Lag 3 Surprise trials, each with a different Surprise stimulus. Six Surprise stimuli were auditory objects drawn from Experiments 1-3: a spoken letter “i”, a spoken digit “one”, an alarm, a cough, a dog barking, and a hiccup. Six other Surprise stimuli were complex tones, each comprised of five log-related frequencies from the following ranges: 455-792, 794-1260, 909-1583, 944-2381, 1349-2142, and 1819-3167 Hz. All stimuli were first equalized by dividing each waveform by its summed absolute amplitude (envelope). This procedure produced sound intensities that we judged to be subjectively similar. Relative to the volume for Surprise stimuli, the probe was set to 45% and distractors were set to 30%. Finally, probe detection was also made slightly easier by the removal of the three highest-pitch distractors.

Before beginning the main experiment, participants completed 8 practice trials at a slower speed (SOA = 130 ms) and 8 practice trials at normal speed (SOA = 120 ms). Practice blocks contained feedback but no Surprise stimuli. For the main experiment, participants completed two blocks of 30 trials each. Participants reported whether the probe was present or absent when prompted at the end of each trial.

Results and Discussion

Four participants (10% of the sample) had responses that were indistinguishable from random responding. Data from these individuals were removed, leaving 35 individuals in the final sample. Given the brevity of the experiment and its similarity to the 20% heterogenous condition in Experiment 3, we focused our analysis on overall Surprise effects. These effects were evidenced by significantly lower performance during Lag 3 Surprise trials (*mean* = 83.49%, *SD* = 27.27) compared to Probe only trials (*mean* = 98.25%, *SD* = 3.37; Wilcoxon signed-ranks test: *p* = .002). In contrast, the hit rate following auditory objects (*mean* = 82.86%, *SD* = 29.16) was similar to the hit rate following complex tones (*mean* = 84.29%, *SD* = 32.70; Wilcoxon signed-ranks test: *p* = 1.00). The false alarm rate was slightly but significantly higher for No probe trials (*mean* = 8.81%, *SD* = 12.77) compared to Surprise only trials (*mean* = 4.76%, *SD* = 16.46; *p* = .04). Importantly, this rate did not differ across auditory object (*mean* = 4.29%, *SD* = 14.20) and complex tone (*mean* = 5.71%, *SD* = 23.55) Surprise stimuli (*p* = 1.00).

Both real-world, recognizable sounds and artificially-generated tones yielded similar

SiD results. This finding is consistent with other auditory experiments involving capture and distraction, in which deviant tones are frequently used to evoke brain responses and induce behavioral costs (Friedman et al., 2001). This experiment also showed that SiD could be induced even when baseline probe detection was virtually perfect.

Experiment 5: Capture in singleton detection mode?

In the first four experiments, we showed that relatively unexpected and rare stimuli can capture attention, thereby impairing probe detection. We argued that the Surprise stimuli disrupted the task despite being irrelevant to searching for the probe item. Another possibility, however, is that participants adopt a “singleton detection mode” to identify the probe item (Bacon & Egeth, 1994; Egeth & Yantis, 1997). As the probe has a significantly higher pitch than the distractor items, participants may simply listen for a deviant sound. If they do so, the Surprise stimulus may capture attention because it, like the probe, is a singleton.

Various features of our experiments suggest that participants were not using “singleton detection mode”, at least not exclusively. The disruptive effects of a given Surprise stimulus were sensitive to its frequency of appearance and relative novelty, with effects habituating across multiple presentations (Experiments 1-3). In addition, SiD was found when participants searched for a non-singleton letter “X” amongst other letters (Experiment 1B). Nevertheless, we sought a more direct test. Therefore, we examined whether SiD would occur when the probe was defined by a specific intermediate pitch amongst other distractor tones. A singleton strategy would likely not work in such circumstances (Bacon & Egeth, 1994).

Method. The current experiment was identical to the Homogenous 20% condition of Experiment 3 with the following exceptions. Forty participants (25 males, mean age = 29.23 years, age range = 22-51) completed the task. The stimuli were five pure tones of log-related frequencies, specifically 396, 629, 1000, 1587, and 2519 Hz. The 1000 Hz tone served as the probe. The number of tones was decreased so that the task remained feasible for participants. To help them further, they completed 8 practice trials at a slower speed (SOA = 130 ms) and 8 practice trials at normal speed (SOA = 120 ms) before beginning the experimental trials. Practice blocks contained feedback but no Surprise stimuli.

Two types of Surprise stimuli were generated from the tones flanking the probe tone (i.e., 1000 Hz) so that they would not be confused for the probe. The “Low” Surprise stimulus was a tone gliding from 396 to 629 Hz, whereas the “High” Surprise stimulus was a tone gliding from 1587 to 2519 Hz. Participants were randomly assigned to either the Low or High condition, in which they heard only the corresponding Surprise stimulus. Stimulus powers were equalized across all the sounds as in Experiment 4.

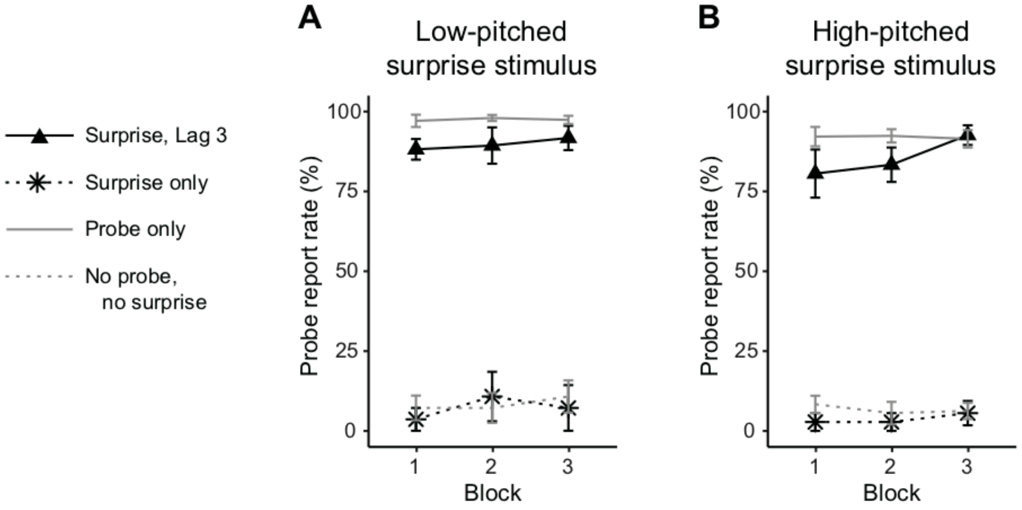


Figure 5. Mean probe report results for Experiment 5 presented by Block and Trial Type for each Surprise Group (A and B). Note that the report rate in the absence of a probe represents false alarms.

Results and Discussion

Participants whose performance was indistinguishable from random responding (n = 8, 20%) were excluded from further analyses. 32 individuals were thus included in the final sample, 14 in the "Low" Surprise group and 18 in the "High" Surprise Group. For visualization purposes, the average probe report rate was calculated by block for each Trial Type and Surprise Group (Figure 5). The data were analyzed using LMEMs with a between-groups variable of Surprise group ("High", "Low"; categorical) and within-subject variables of Time (continuous) and Trial Type (Probe only, Lag 3; categorical). For hits, there was a significant main effect of Trial Type ($\chi^2(1) = 21.22$, $p < .001$), a significant Time x Trial Type interaction ($\chi^2(1) = 8.24$, $p = .004$), and a significant Trial Type x Surprise Group interaction ($\chi^2(1) = 4.10$, $p = .043$). The main effect reflected reduced performance for Lag 3 Surprise trials, whereas the significant interaction was due to SiD habituation. Across Time, Surprise trial performance improved ($p = .005$), whereas Probe only performance did not significantly change ($p = 1.00$). Follow-up Wilcoxon signed-rank tests showed that Lag 3 Surprise performance was lower only during the first block ($p = .007$; subsequent blocks: *p*s > .06). Indeed, the hit rate for the very first Surprise trial (21 of 32; 66%) was marginally lower than the hit rate for the preceding Probe only trial (28 of 32, 88%; proportion test: $\chi^2(1) = 3.13$, $p = .076$). No other main effects or interactions were significant for hits (χ^2s < 1.82, *p*s > .18). For false alarms, there were no significant main effects or interactions (χ^2s < 1.89, *p*s > .17).

This experiment shows that Surprise-induced Deafness, both the deficit itself and its characteristic pattern of habituation, is still present when the probe is not a singleton amongst the distractor items. We conclude that SiD depends critically on stimulus-driven attentional capture by a salient task-irrelevant item. Other factors, including goal-directed attentional influences, may contribute to the deficit as well. We explore this idea in a final pair of experiments.

Experiment 6: Replicating and extending the Auditory Attentional Blink (AAB)

In our Surprise-induced Deafness paradigms, participants search for a probe item amongst distractors, and failure to detect that probe is the critical measure. These aspects are also central to the auditory attentional blink (AAB) (Arnell & Jolicoeur, 1999; Mondor, 1998). Despite their general similarity, the two paradigms have a crucial difference: Whereas SiD is caused by an unexpected and task-irrelevant stimulus, the AAB is caused by processing a target item that precedes the probe. The AAB substantially—sometimes completely—attenuates when participants are told to ignore the target stimulus (Arnell & Jolicoeur, 1999; Horváth & Burgyán, 2011). As such, the AAB has been argued to be due to goal-directed attentional processing of the target, not stimulus-driven attentional capture (Arnell & Jenkins, 2004; Arnell & Jolicoeur, 1999; Martens et al., 2009; Martens, Kandula, & Duncan, 2010; Martens, Wierda, Dun, de Vries, & Smid, 2015; Mondor, 1998; Shen & Mondor, 2006; Soto-Faraco & Spence, 2002; Vachon & Tremblay, 2005, 2006; Vachon, Tremblay, Hughes, & Jones, 2009).

In the current experiment, we tested whether an AAB would be found with a paradigm adapted from our SiD experiments. This approach allowed us to compare the two deficits when using highly similar procedures and stimuli. For example, three complex tones that were used as Surprise stimuli in Experiment 4 were here used as targets. We could therefore test the effects of the same stimuli when they played different roles. The experiment also allowed us to prepare for Experiment 7, in which we used an individual differences approach to further investigate the relationship of SiD and AAB.

Method

Participants. Forty-three NUS undergraduates with normal hearing participated for course credit. We used a target-probe AAB design for similarity to our SiD paradigm. Participants were randomly allocated to either an Experimental group (detect and report both the target and the probe) or to a Control group (detect and report only the probe). Three participants in the former group (age and gender not recorded) found the task too difficult during the practice trials and elected not to continue. Each group therefore had 20 individuals (Experimental: 6 males, mean age = 20.30, age range = 18-24; Control: 9 males, mean age = 21.7, age range = 19-25).

Stimuli and apparatus. Distractor stimuli were pure tones of log-related frequencies ranging from 639 to 2911 Hz (Figure 6A). The same stimulus set (save 3175 Hz) had been used for Experiments 1-3. The probe was the same 4000 Hz pure tone as before. Three target stimuli were constructed from five log-related frequencies: 455-792 Hz ("Low"), 909-1583 Hz ("Middle"), and 1819-3167 Hz ("High"). Note that these stimuli were used as "Surprise" stimuli in Experiment 4, thereby allowing us to compare the effects of the same stimuli playing different roles. All stimuli were equalized by dividing each waveform by its summed absolute amplitude (envelope). The experiment was conducted in the laboratory, with PsychoPy software (Peirce, 2007) and a Dell computer (OPTIPLEX 990) used for stimulus presentation and data collection. Sounds were presented binaurally through TDK headphones (ST 100) with an approximate maximum volume of 70 dB. Participants made their responses through a standard computer keyboard.

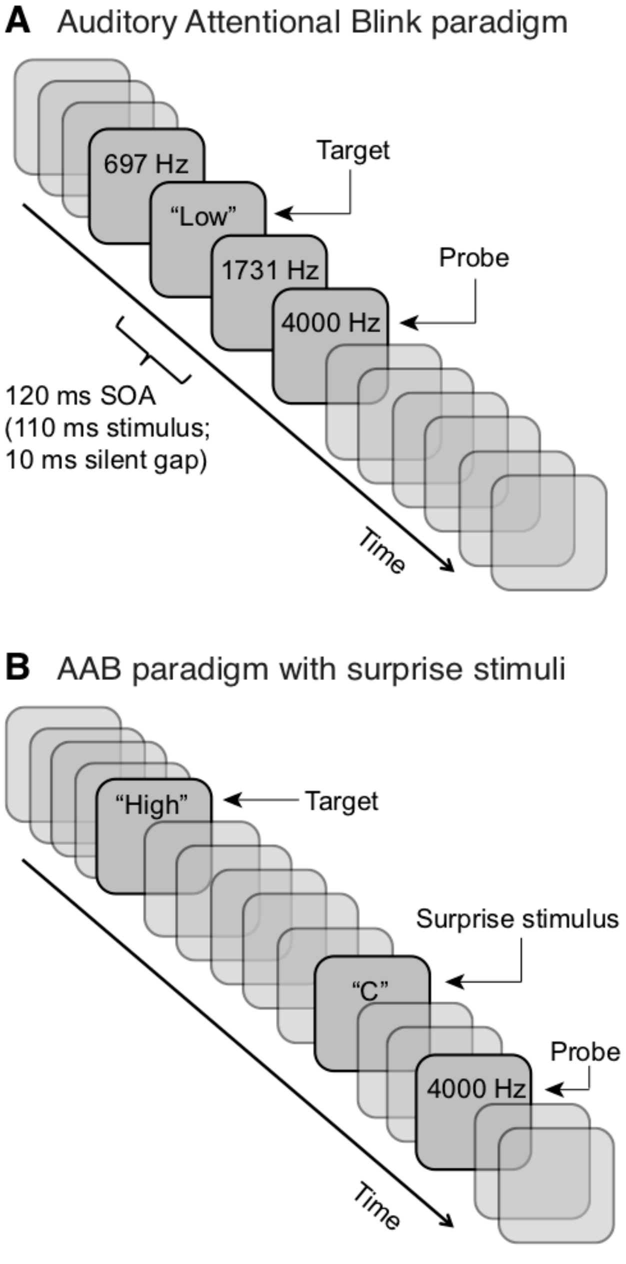


Figure 6. A) Stimulus sequence for trials in the Auditory Attentional Blink (AAB) paradigm in Experiment 6. A Lag 2 trial is shown. B) Stimulus sequence for trials in the AAB-SiD hybrid paradigm in Experiment 7. A Lag 3 Surprise trial is shown. Note that Surprise trials with the probe were also Lag 9 AAB trials.

Task procedures. Before each trial, the task-relevant tones (target and probe for the Experimental group, only the probe for the Control group) were played to refresh participants' auditory memory. Participants could play these demo tones as many times as they chose. A fixation cross was then displayed until the participant pressed any key to begin a rapid auditory presentation (RAP) of 30 tones. One of the three targets was presented during each trial, followed by the probe on 75% of trials. When present, the probe sounded at Lag 1, 2, 3, 5, or 8 relative to the target. The probe was presented after between 17 and 27 items on every trial. Participants reported the probe as soon as it was detected, thereby giving us a secondary measure of performance impairment, as in Experiment 1. If participants did not make a response during the RAP, they were prompted by a question ("Probe?") 300 ms after the RAP's conclusion. Participants pressed the '<' key for "present" and the '>' key for "absent". Participants in the Experimental group were then prompted to indicate which target was presented ("Target?") by pressing the '1', '2', or '3' key. For probe detection, both accuracy and speed were emphasized ("Respond as quickly as it is possible (but not at the expense of accuracy) using your right hand"), whereas speed was not emphasized for target discrimination.

Participants completed 240 trials presented in 6 blocks; each block contained 6 trials for each target-probe Lag and 10 No probe trials. Participants were encouraged to take brief breaks between blocks. Before beginning the main task, participants in the Experimental group practiced at least three phases of 6 trials each: target discrimination, probe detection, and both tasks (Lag 8 only). Participants in the Control group practiced only probe detection. For each phase of practice, participants had to reach an accuracy of 66% before advancing to the next phase.

Statistical approach. For the auditory attentional blink in Experiments 6 and 7, the key variables of were all categorical. Therefore, we assessed their effects on response rates and times using standard repeated-measures ANOVAs using ezANOVA (version 4.4-0) (Lawrence, 2016). Greenhouse–Geisser correction was applied when the assumption of sphericity was found to have been violated. For such cases and when variances were unequal in independent-samples t-tests, we report uncorrected degrees of freedom and corrected F-values and p-values. Effect sizes for significant results are reported as generalized eta-squared values (denoted as "η_G^2"), which are presented after the corresponding p-values.

Results and Discussion

No participants were removed for low probe detection performance based on the criteria described in Experiment 1. Two participants, however, had target discrimination accuracy that was not distinguishable from chance performance (<40%). These individuals were removed from the sample, leaving 18 in the Experimental group.

Target discrimination. The average report accuracy was 69.86% (*SD* = 12.61%), indicating a challenging task. This rate did not significant vary by Lag (repeated-measures ANOVA: *F*(4, 68) = 1.18, *p* = .33). Furthermore, target performance did not significantly differ across trials with a probe (*mean* = 69.20, *SD* = 12.94) compared to those without one (*mean* = 71.85, *SD* = 13.08; paired-samples t-test: *t*(17) = 1.58, *p* = .13).

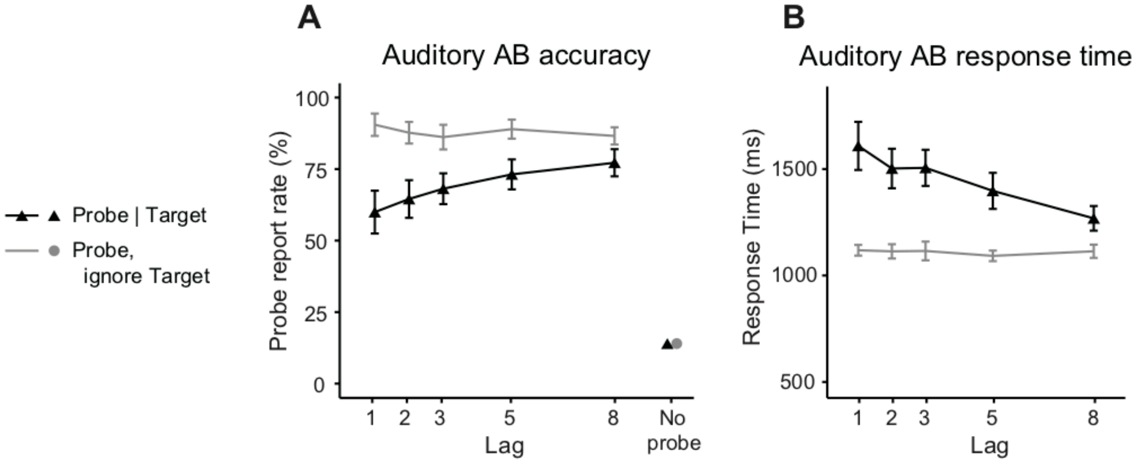


Figure 7. Probe performance results for Experiment 6. A) Mean probe report by Lag for the

Experimental and Control groups. For the former, probe report was calculated contingent on correct target identification (Probe|Target). The report rate for "No probe" represents the false alarm rate. B) Mean probe response times by Lag for the Experimental and Control groups. Trials for the former were again contingent on correct target identification.

Probe report. For the Experimental group, report rates were calculated from trials with correct target report. Mean report rates were submitted to a mixed 5 × 2 two-way ANOVA with Lag as a within-subject variable and Group as a between-subject variable. Both the main effect of Lag ($F(4, 144) = 3.31, p = .013, \eta_G^2 = .015$) and Group ($F(1, 36) = 9.51, p = .004, \eta_G^2 = .18$) were significant, as was their interaction ($F(4, 144) = 6.60, p < .001, \eta_G^2 = .03$). Performance improved with increasing Lag for the Experimental group (repeated-measures ANOVA: ($F(4, 68) = 5.43, p = .01, \eta_G^2 = .06$), but not for the Control group ($F(4, 76) = 1.84, p = .15$). Compared to the Control group, performance in the Experimental group was significantly lower for Lags 1 and 2 (independent samples t-tests: ps $< .026$), marginally lower for Lags 3 and 5 (ps $< .067$), and not significantly different for Lag 8 ($p = .48$). Finally, the mean false alarm rates did not significantly differ across Groups ($t(36) = 0.04, p = .97$).

These results are consistent with the presence of an auditory attentional blink. The effect lasted for roughly half a second within each trial, in line with previous AAB reports (Mondor, 1998; Vachon & Tremblay, 2005) and similar to the timecourse of SiD. Unlike in SiD, however, goal-directed processing of a target caused the AAB deficit. Participants in the Control group, who ignored target stimuli, evidenced no lag-dependent probe detection deficits. These same target stimuli could induce deficits through stimulus-driven attentional capture when they were relatively unexpected and rare, as evidenced in Experiment 4.

Probe reaction times. For the Experimental group, only trials with correct target reports were included for the analysis. Mean reaction times were submitted to a mixed 5 × 2 ANOVA with Lag as a within-subject variable and Group as a between-subject variable. Both main effects (Lag, $F(4, 144) = 3.95$, $p = .015$, $\eta_G^2 = .025$; Group, ($F(1, 36) = 18.92$, $p < .001$, $\eta_G^2 = .29$)) and their interaction ($F(4, 144) = 6.60$, $p < .001$, $\eta_G^2 = .03$) were significant. Reaction times significantly decreased with increasing Lag for the Experimental group (repeated measures ANOVA: $F(4, 68) = 5.65$, $p = .005$, $\eta_G^2 = .08$), but not for the Control group ($F(4, 76) = 1.12$, $p = .34$). Reaction times for the Experimental group were significantly higher than for the Control group at Lags 1-5 (independent samples t-tests: ps $< .013$) but not at Lag 8 ($p = .35$). Although not necessarily indicative of an auditory attentional blink *per se* (Arnell & Jolicoeur, 1999; Ruthruff & Pashler, 2001), these reaction time results reflect a cost to attending to the first target. Consistent with the probe hit rate results presented above, the first target only imposed this cost when it was task-relevant.

Experiment 7: The relationship between SiD and the AAB

In Experiment 6, we successfully demonstrated an auditory attentional blink (AAB) using a paradigm adapted from our SiD experiments. In this final experiment, we compared the two effects directly. Despite their similarity in procedure and results, SiD and the AAB are associated with different failures of attentional control. SiD is induced by a Surprise stimulus redirecting attention in a stimulus-driven fashion, whereas the AAB is caused by goal-directed attentional processing of a target item. Our experiments clearly reflect this distinction. In our paradigm, an AAB was not triggered by stimulus-driven attentional capture, whereas SiD was found even for the very first Surprise trial with stimuli that contained no target-defining features.

Nevertheless, there are reasons to believe that the two deficits share at least some common mechanisms. Foremost, it is not clear whether goal-directed and stimulus-driven attentional control processes themselves are fully dissociable. Although some researchers have argued that the two sources of attentional control are independent (Pinto, van der Leij, Sligte, Lamme, & Scholte, 2013), others have argued that they cannot function in isolation (Rauschenberger, 2010) or that they involve the same neural circuitry (Buschman & Miller, 2007). The consensus view appears to stake out a middle position: Goal-directed and stimulus-driven attention are partially dissociable neurally and psychologically (Asplund, Todd, Snyder, & Marois, 2010; Corbetta & Shulman, 2002; Egeth & Yantis, 1997; Serences et al., 2005). Such tentative conclusions hold for both visual and auditory attentional control (Alho, Salmi, Koistinen, Salonen, & Rinne, 2015). For example, task-irrelevant sounds elicit a novelty-P3 electrophysiological response, which differs from the target-P3 in its habituation, more rapid onset, and more frontal scalp distribution (Debener et al., 2002). Such results suggest largely dissociable attentional forms. Conversely, attentional orienting caused by goal-directed or stimulus-driven factors recruits many of the same brain regions, including temporo-parietal, superior parietal, and frontal cortex (Alho et al., 2015; Salmi, Rinne, Koistinen, Salonen, & Alho, 2009). Regardless of the degree of overlap, the two forms of control interact during normal function of either visual or auditory attention (Asplund, Todd, Snyder, & Marois, 2010; Bidet-Caulet, Bottemanne, Fonteneau, Giard, & Bertrand, 2015; Corbetta & Shulman, 2002; Serences et al., 2005; Sussman, Winkler, & Schröger, 2003).

SiD and the AAB might reflect a common underlying processing limitation even if each involves a different and dissociable form of attentional control. For example, in many paradigms, attention must be spatially redirected. A spatial shift could involve common neuropsychological processes even if its triggers are distinct. Importantly, this potential common process is unlikely to underlie SiD and the AAB since neither involves spatial redirection. The deficits converge elsewhere, however: Each fundamentally involves the disruption of a goal-directed task. Indeed, Asplund et al. (2010) suggested that the point of overlap between Surprise-related activation and goal-directed attentional processing could be where the attentional forms interact (Asplund, Todd, Snyder, & Marois, 2010).

In the current experiment, we used an individual differences approach to better understand the relationship between SiD and the AAB. If the deficit magnitudes are correlated across participants, the inference is that they rely on shared cognitive mechanisms. This approach and reasoning have been used to show that attentional blinks draw on separate resources across modalities (Martens et al., 2009, 2010), and that different types of visual ABs (broadly, those with probe detection versus target discrimination for the second item) have partially dissociable causes (Dale, Dux, & Arnell, 2013; Kelly & Dux, 2011). To keep the experimental session manageably brief, we estimated the two deficits' magnitudes within a single paradigm. In addition, we included a relatively small Control group who ignored the targets (see Experiment 6). This manipulation ensured that the targets did not produce probe detection deficits in our specific paradigm.

Method

The current experiment combined the AAB and SiD paradigms, adapting methods from Experiment 6 and Experiment 3 (20% Varying condition) with additional modifications. For clarity, the key design elements are reproduced below.

Participants. Two hundred participants (115 males, 4 gender not reported; mean age = 30.74 years, age range = 20-61, 2 age not reported) completed the experiment on Amazon Mechanical Turk (AMT). This sample size allowed us to detect small correlation effects (r = 0.2) with 80% power, which we deemed necessary given that even different forms of the visual attention blink paradigm correlate at only a moderate level (r = 0.4) (Dale et al., 2013; Kelly & Dux, 2011; MacLean & Arnell, 2012). Due to a data collection error, the data from one participant could not be used. In addition to this Experimental group, 25 participants (13 males, mean age = 30.28 years, age range = 24-39) completed a Control version of the experiment, in which the target was ignored. This smaller Control group allowed us to test whether the targets captured attention in a stimulus-driven fashion, and whether targets affected SiD itself. All participants received $3 USD for the study, which lasted approximately 50-60 minutes.

Stimuli and apparatus. As in Experiment 6, distractor stimuli were pure tones of log-related frequencies ranging from 639 to 2911 Hz, and the probe was a 4000 Hz pure tone. The two complex tone targets were constructed from five log-related pure tones, with ranges of 794-1260 and 1349-2142 Hz. The number of target alternatives was reduced in an attempt to shorten the blink period, thereby more clearly separating within-blink probe detection performance from probe detection that was not affected by target processing (Shore, McLaughlin, & Klein, 2001). Twenty-four Surprise stimuli from Experiments 1-5 were selected for the present experiment. The set consisted of spoken letters (‘C’, ‘L’, ‘J’, ‘H’), spoken digits (‘one’, ‘two’), and environmental sounds (alarm, balloon popping, car honking, cat meowing, cough, cowbell, dog barking, giggle, hiccup, hi-hat cymbal, light bulb breaking, mosquito buzzing, plunger, popping tongue, slide whistle, snare drum, sneeze, slurping). Widely varying Surprise stimuli were used in an effort to reduce SiD habituation, enabling us to obtain more robust effects and stable individual differences. As in Experiment 6, all stimuli were first equalized by dividing each waveform by its summed absolute amplitude (envelope). They were then adjusted so that their amplitudes were 30% (distractors and Surprise stimuli) or 45% (probe) of the targets’ amplitudes.

Task procedure. Before each trial, the target and probe stimuli were played in sequence to refresh participants’ auditory memory. Participants could play these demo tones as many times as they chose. A fixation cross was then displayed until the participant pressed any key to begin a rapid auditory presentation (RAP) of 30 tones. One of the two targets was presented during each trial, which was followed by the probe on 75% of trials. When present, the probe sounded at Lag 1, 2, 3, 5, 7, or 9 relative to the target. The probe was presented after 17 to 27 RAP items. During 20% of the trials, a Surprise stimulus was played six items (720 ms) after the target (Figure 5B). At Lag 6, the Surprise stimulus would be largely clear of the AAB window; pilot testing confirmed that the Surprise stimulus was clearly audible (and reported to be distracting) in this position. 75% of the Surprise trials contained a probe, which followed the Surprise stimulus at Lag 3. This timing rendered the probe vulnerable to SiD, while putting it outside the AAB window (Lag 9). At the conclusion of each RAP, participants were asked to identify the target by pressing ‘d’ or ‘f’. They were next asked about the probe, responding with ‘j’ for present and ‘k’ for absent. All responses were unspeeded.

Participants completed 120 trials presented in 3 blocks. Each block contained exactly 6 Surprise trials with a probe, 2 Surprise trials without a probe, 24 standard trials with a probe (4 per Lag), and 8 standard trials without a probe. The trials were presented in a different random order for each block. Importantly, each participant experienced the same trial order, thereby reducing performance variability due to trial/stimulus differences. Before beginning the main task, participants were instructed to adjust the volume to their level of comfort and to use headphones if possible. They then completed 16 practice trials with feedback: 4 for target discrimination, 4 for probe detection, and 8 during which both tasks were performed. Practice trials included only the relevant stimuli (e.g. no targets during probe detection practice). No Surprise stimuli were presented during the practice block.

The Control group’s version of the experiment was procedurally identical with the following exceptions. Participants were explicitly instructed to ignore the targets, and they were not queried about them at the conclusion of each trial. Only the probe sound was played before each trial, and only four practice trials (for probe detection) were completed.

Analytical approach. Task performance for the SiD and AAB paradigm components was first assessed using the approaches from previous experiments. Data from participants with low probe detection or target discrimination performance removed as before. Conversely, preliminary analyses also revealed that some participants evidenced probe detection that was consistently high, perhaps indicative of ceiling effects. As these individuals provided little behavioral variance of use, we removed their data for our correlation analyses only. Participants were classified in this group if their probe detection hit rates, contingent upon correct target discrimination (Probe|Target), were above 90% for both short lags (average of Lags 1 and 2) and long lags (average of Lags 7 and 9). Critically, we examined and report the correlations with these individuals included in the sample as well.

For correlation analyses, the magnitude of each effect for each participant was calculated using regression (MacLean & Arnell, 2012). This approach controls for variation in baseline performance across participants. For SiD, we regressed probe detection hit rates from Lag 9 trials against hit rates for Surprise trials that contained a probe. The residuals were then saved and used for subsequent analyses (Dale et al., 2013). For the AAB, we regressed probe detection hit rates at long lags (Lags 7 and 9) against those at short lags (Lags 1 and 2), again saving the residual variability. For the AAB residuals, the input probe performance was calculated contingent on correct target identification; for SiD residuals, it was not. The metrics from which each set of residuals were calculated were chosen because they best reflected the deficits in question based on group-average performance and previous findings (Arnell & Jolicoeur, 1999; Asplund, Todd, Snyder, Gilbert, & Marois, 2010; Shen & Mondor, 2008). For completeness, however, we considered two additional residual calculations. First, to compare the deficits at the same Lag, we calculated AAB residuals by regressing probe detection at long lags (Lags 7 and 9) against Lag 3. Second, to compare the deficits by controlling for the same baseline performance, we calculated SiD residuals by regressing probe detection at long lags (Lags 7 and 9, contingent on correct target discrimination) against Surprise trials that contained a probe.

The residuals for each deficit were used in two different ways. First, we calculated internal-consistency reliability scores of our effect measures using a split-half approach. The relevant trials for each metric component (e.g. Lag 7 and 9 trials, contingent on correct target discrimination) were identified and then split into odd or even trials according to their order of appearance. Residuals were calculated from each set of metric components, after which they were correlated across odd and even trials. Second, we calculated Pearson correlations amongst target discrimination accuracy, probe detection hit rates (not contingent on correct target discrimination), SiD magnitude (residuals), and AAB magnitude (residuals).

Results and Discussion

In the Experimental group, 55 participants (28% of the sample) had probe detection performance below d'=0.5. An additional 14 participants (10% of the remaining sample) had target discrimination accuracy below 60%, which was indistinguishable from chance. The data from these participants were not included in subsequent analyses, leaving 130 participants in the final sample. Five participants (20% of the sample) in the Control group likewise had low probe detection performance. Their data were removed, leaving 20 participants in the final sample.

SiD: Probe reports. In addition to the Surprise trials themselves, standard Lag 9 and No probe trials were included for the SiD analyses. The former provided a baseline for probe hit rates, whereas the latter provided the same for false alarms. For visualization, the mean hit and false alarm rates were calculated separately for each block (Figure 8A). For the Experimental group, the main effect of Trial Type for hits was significant ($\chi^2(1) = 46.20$, $p < .001$), reflecting lower performance for Surprise trials. Neither the main effect of Time nor the Time x Trial Type interaction were significant (*p*s > .19). Follow up tests found that the Surprise trial hit rate was lower than Probe only performance in each block (Wilcoxon signed ranks tests, $p < .001$). For false alarms, there was a marginally significant main effect of Time x Trial Type interaction ($\chi^2(1) = 3.07$, $p = .080$). No main effects were significant (ps > .14). Consistent with the findings in Experiments 1-5, the Surprise stimuli induced a probe detection deficit, which gradually—but far from completely—habituated across trials. The Surprise results for the Control group were similar (data not shown).

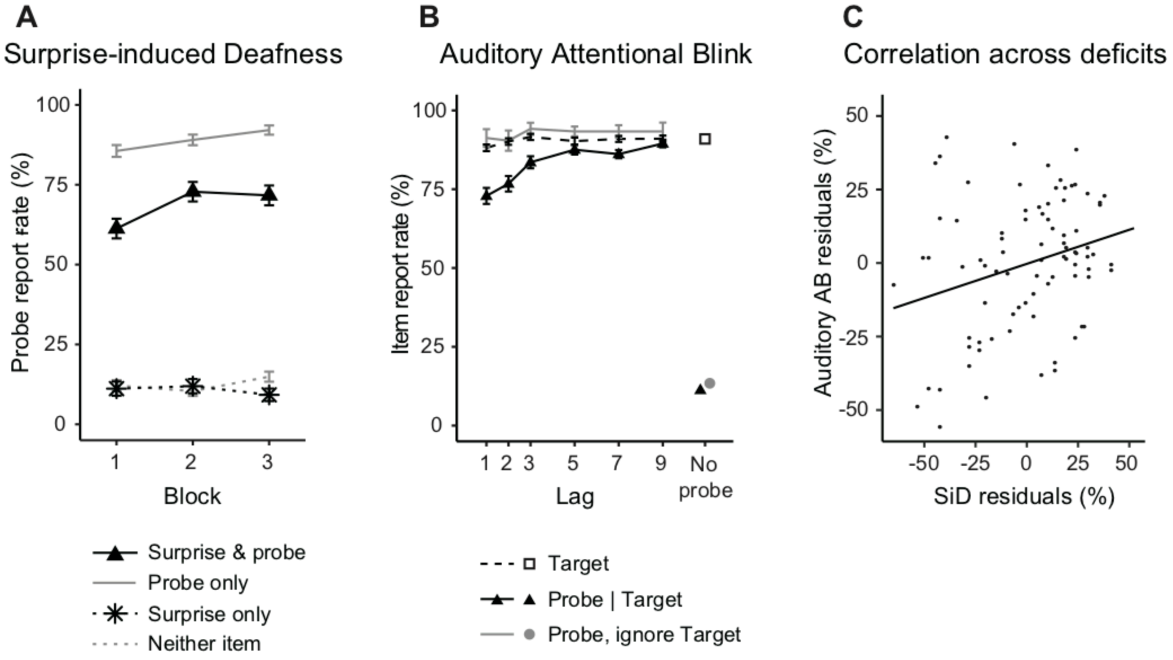


Figure 8. Performance in Experiment 7. A) Mean target detection rate for each Trial Type as a function of block for the Experimental group. Note that all trials also contained a target, though performance was not calculated contingent on correct target discrimination. “Probe only” trials refer to Lag 9 trials, and “Neither item” trials refer to target only trials. B) Mean probe detection rate contingent on correct target discrimination as a function of Lag. Data for the “Probe, ignore target” condition are from the Control group, whereas other results are from the Experimental group. C) Scatterplot of SiD and AAB residuals, showing the line of best fit. Error bars for panels A and B represent ±1 standard error of the mean (SEM).

AAB: Target discrimination. Correct target discrimination was 90.06% (*SD* = 9.42) overall, showing that the task was manageable for most participants. Target discrimination also varied slightly but significantly by Lag (Figure 8B; repeated-measures ANOVA: $F(5, 645) = 2.94$, $p = .01$, $\eta_G^2 = .009$). Follow-up analysis revealed that the Target discrimination rate at Lag 1 was significantly lower than at Lags 3, 7, and 9 (*p*s < .036) but not significantly different from Lags 2 and 5 (*p*s > .19). Similar effects of Lag on target performance have been reported in other AAB studies (Martens et al., 2009; Shen & Alain, 2012).

AAB: Probe reports. Mean probe performance was calculated for each Lag, conditional on correct target discrimination for the Experimental group (Probe|Target; Figure 8B). Mean hit rates by participant were submitted to a mixed 6 × 2 ANOVA with Lag as a within-subject variable and Group (Experimental, Control) as a between-subject variable. This analysis revealed significant main effects of Group ($F(1, 148) = 5.82$, $p = .017$) and Lag ($F(5, 740) = 32.25$, $p < .001$), as well as a significant interaction ($F(5, 740) = 11.70$, $p = .023$). Performance improved with increasing Lag for the Experimental group (repeated-measures ANOVA: $F(5, 645) = 32.94$, $p < .001$, $\eta_G^2 = .071$), but not for the Control group ($F(5, 95) = 0.51$, $p = .71$). Compared to the Control group, performance in the Experimental group was significantly lower for Lags 1, 2, 3, and 7 (independent samples t-tests: *p*s < .03), marginally lower for Lag 5 ($p = .06$), and not significantly different at Lag 9 ($p = 1$). Finally, the false alarm rates between two groups did not significantly differ between the two groups ($t(148) = -0.42$, $p = .67$). Overall, the results demonstrate a convincing AAB, with a lag-dependent deficit that only appeared when the target was attended (Figure 8B).

To examine the possibility that target discrimination affected SiD, we calculated the magnitude of the deficit for each participant in Experimental and Control groups. The probe hit rate during Surprise trials was subtracted from this rate during standard Lag 9 trials. SiD magnitude for the Experimental group (*mean* = 0.20, *SD* = 0.25) was not significantly different from that of Control group (*mean* = 0.14, *SD* = 0.23; *t*(148) = 1.13, *p* = .26). We also compared the Experimental group's SiD magnitude to the Varying 20% condition of Experiment 3 (*mean* = 0.20, *SD* = 0.19). There was no significant difference between the two rates (*t*(198) = 0.01, *p* = .99). We conclude that target discrimination did not interfere with perception of the Surprise stimulus or its subsequent effects on probe detection.

Preparation for correlation analyses. A subset of the participant sample (46 of 130; 36%) had nearly perfect probe detection performance regardless of the condition. These individuals were removed for our primary correlation analyses, leaving 84 in the sample to be analyzed. As noted above, for completeness we also analyzed the full sample of 130 individuals and also report those results.

Internal-consistency reliability. Metric stability was high for both SiD (split-half *r*(82) = .74, *p* < .001) and the AAB (split-half *r*(82) = .65, p < .001). These values represent high reliability and are comparable to those for the visual AB (Dale et al., 2013; Kelly & Dux, 2011). Note that we did not correct our coefficients using the Spearman-Brown procedure (Spearman, 2010) because far more AAB-relevant trials contribute to the cross-deficit correlations that are of primary interest; therefore, the uncorrected AAB reliability values likely represent a sensible floor. When SiD residuals were instead calculated using Lag 7 and 9 trials as the baseline, similar reliability was obtained (*r*(82) = .73, *p* < .001). When AAB residuals were calculated from Lag 3 trials, the reliability was lower but still significant (*r*(82) = .33, *p* = .010). The lower reliability is likely due to smaller overall AAB effects at Lag 3, though it is worth noting that the residuals from Lag 3 and the residuals from Lags 1 and 2 were highly correlated (*r*(82) = .62, *p* < .001). Finally, when the residuals were calculated from the complete sample, the reliability scores were again similar for both SiD (*r*(128) = .67, *p* < .001) and the AAB (*r*(128) = .60, *p* < .001).

Correlation between SiD and AAB deficits. To test the relationship between SiD and the AAB, we calculated the correlation between the deficit magnitudes. A relatively small but significant correlation was found (Table 6). When SiD residuals were instead calculated using Lag 7 and 9 trials as the baseline, the SiD-AAB correlation was similar ($r(82) = .29$, $p = .008$). The same was true when AAB residuals were calculated from Lag 3 trials ($r(82) = .28$, $p = .010$), despite the much lower AAB reliability at Lag 3. Finally, when the residuals were calculated from the complete sample, the correlation was again comparable ($r(128) = .38$, $p < .001$). Across different residual calculations, the correlation coefficient for SiD and the AAB indicated a small-to-moderate effect size.

Table 6. Correlation coefficients (*r*-values) for different metrics considered pairwise in Experiment 7. The lack of correlation between probe detection and the deficit residuals is expected because baseline probe detection performance was statistically removed. * $p = .014$, ** $p < .001$.

Measure	1	2	3
1. Target discrimination	–		
2. Probe detection (hit rate)	.38**	–	
3. SiD magnitude	-.09	.00	–
4. AAB magnitude	-.16	.05	.27*

The magnitude of the SiD-AAB correlation was found to be far lower than the reliability metrics for each deficit, suggesting a partial dissociation between the processing limitations behind SiD and the AAB. This result is consistent with the partial dissociation between stimulus-driven and goal-directed attentional control that enjoys broad empirical support (Alho et al., 2015; Corbetta & Shulman, 2002; Egeth & Yantis, 1997). We further consider the meaning of both the partial overlap and partial dissociation in the general discussion.

The present experiment contained a number of limitations that are also worth noting. The experiment was conducted in a single session online, precluding the collection of test-retest measures across days. Such measures would be more compelling evidence of stable individual differences. Furthermore, the deficits were studied within the same paradigm, meaning that the SiD results were obtained from participants engaged in a dual-task situation (target discrimination followed by probe detection). We have addressed these limitations in a separate study in our laboratory, in which we found similar correlation magnitudes between SiD and the AAB, as well as their visual counterparts (Liaw, Obana, Chia, & Asplund, in prep.).

References

Alho, K., Salmi, J., Koistinen, S., Salonen, O., & Rinne, T. (2015). Top-down controlled and bottom-up triggered orienting of auditory attention to pitch activate overlapping brain networks. *Brain Research*, *1626*, 136–145. https://doi.org/10.1016/j.brainres.2014.12.050

Arnell, K. M., & Jenkins, R. (2004). Revisiting within-modality and cross-modality attentional blinks: Effects of target—distractor similarity. *Perception & Psychophysics*, *66*(7), 1147–1161. https://doi.org/10.3758/BF03196842

Arnell, K. M., & Jolicoeur, P. (1999). The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology. Human Perception and Performance*, *25*(3), 630–648.

Asplund, C. L., Todd, J. J., Snyder, A. P., Gilbert, C. M., & Marois, R. (2010). Surprise-induced blindness: A stimulus-driven attentional limit to conscious perception. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1372–1381. https://doi.org/10.1037/a0020551

Asplund, C. L., Todd, J. J., Snyder, A. P., & Marois, R. (2010). A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention. *Nature Neuroscience*, *13*(4), 507–512. https://doi.org/10.1038/nn.2509

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. https://doi.org/10.1016/j.tics.2012.06.010

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, *55*(5), 485–496. https://doi.org/10.3758/BF03205306

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Belopolsky, A. V., Schreij, D., & Theeuwes, J. (2010). What is top-down about contingent capture? *Attention, Perception, & Psychophysics*, *72*(2), 326–341. https://doi.org/10.3758/APP.72.2.326

Bidet-Caulet, A., Bottemanne, L., Fonteneau, C., Giard, M.-H., & Bertrand, O. (2015). Brain Dynamics of Distractibility: Interaction Between Top-Down and Bottom-Up Mechanisms of Auditory Attention. *Brain Topography*, *28*(3), 423–436. https://doi.org/10.1007/s10548-014-0354-x

Brosch, M., & Schreiner, C. E. (1997). Time Course of Forward Masking Tuning Curves in Cat Primary Auditory Cortex. *Journal of Neurophysiology*, *77*(2), 923–943.

https://doi.org/10.1152/jn.1997.77.2.923

Buschman, T. J., & Miller, E. K. (2007). Top-down versus Bottom-up Control of Attention in the Prefrontal and Posterior Parietal Cortices. *Science, New Series*, *315*(5820), 1860–1862.

Chun, M. M., & Marois, R. (2002). The dark side of visual attention. *Current Opinion in Neurobiology*, *12*(2), 184–189. https://doi.org/10.1016/S0959-4388(02)00309-4

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, *39*(2), 131–143. https://doi.org/10.1016/0013-4694(75)90003-6

Crowder, R. G. (1993). Auditory memory. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 113–145). Oxford University Press.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Dale, G., Dux, P. E., & Arnell, K. M. (2013). Individual differences within and across attentional blink tasks revisited. *Attention, Perception, & Psychophysics*, *75*(3), 456–467. https://doi.org/10.3758/s13414-012-0415-8

Dalton, P., & Lavie, N. (2004). Auditory Attentional Capture: Effects of Singleton Distractor Sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(1), 180–193. https://doi.org/10.1037/0096-1523.30.1.180

Debener, S., Kranczioch, C., Herrmann, C. S., & Engel, A. K. (2002). Auditory novelty oddball allows reliable distinction of top–down and bottom–up processes of attention. *International Journal of Psychophysiology*, *46*(1), 77–84. https://doi.org/10.1016/S0167-8760(02)00072-7

Delahaye, R., Fantini, D. A., & Meddis, R. (1999). Effects of practice on performance for different masking tasks. *The Journal of the Acoustical Society of America*, *105*(2), 1153–1153. https://doi.org/10.1121/1.425482

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2002). A Cortical Network Sensitive to Stimulus Salience in a Neutral Behavioral Context Across Multiple Sensory Modalities. *Journal of Neurophysiology*, *87*(1), 615–620. https://doi.org/10.1152/jn.00636.2001

Dux, P. E., & Marois, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception & Psychophysics*, *71*(8), 1683–1700. https://doi.org/10.3758/APP.71.8.1683

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*(1), 269–297.

Escera, C., Alho, K., Schröger, E., & Winkler, I. W. (2000). Involuntary Attention and Distractibility as Evaluated with Event-Related Brain Potentials. *Audiology and Neurotology*, *5*(3–4), 151–166. https://doi.org/10.1159/000013877

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural Mechanisms of Involuntary Attention to Acoustic Novelty and Change. *Journal of Cognitive Neuroscience*, *10*(5), 590–604. https://doi.org/10.1162/089892998562997

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology. Human Perception and Performance*, *18*(4), 1030–1044.

Font, F., Roma, G., & Serra, X. (2013). *Freesound technical demo*. 411–412. https://doi.org/10.1145/2502081.2502245

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression*. SAGE. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain’s evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, *25*(4), 355–373. https://doi.org/10.1016/S0149-7634(01)00019-7

Friedman, D., & Simpson, G. V. (1994). ERP amplitude and scalp distribution to target and novel events: Effects of temporal order in young, middle-aged and older adults. *Cognitive Brain Research*, *2*(1), 49–63. https://doi.org/10.1016/0926-6410(94)90020-5

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention—Focusing the searchlight on sound. *Current Opinion in Neurobiology*, *17*(4), 437–455. https://doi.org/10.1016/j.conb.2007.07.011

Grysman, A. (2015). Collecting Narrative Data on Amazon's Mechanical Turk: Narrative data on Amazon's Mechanical Turk. *Applied Cognitive Psychology*, *29*(4), 573–583. https://doi.org/10.1002/acp.3140

Horstmann, G. (2005). Attentional Capture by an Unannounced Color Singleton Depends on Expectation Discrepancy. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 1039–1060. https://doi.org/10.1037/0096-1523.31.5.1039

Horstmann, G. (2015). The surprise-attention link: A review: The surprise-attention link. *Annals of the New York Academy of Sciences*, *1339*(1), 106–115. https://doi.org/10.1111/nyas.12679

Horváth, J., & Burgyán, A. (2011). Distraction and the auditory attentional blink. *Attention, Perception, & Psychophysics*, *73*(3), 695–701. https://doi.org/10.3758/s13414-010-0077-3

Horváth, J., & Winkler, I. (2010). Distraction in a continuous-stimulation detection task. *Biological Psychology*, *83*(3), 229–238.

https://doi.org/10.1016/j.biopsycho.2010.01.004

Horváth, J., Winkler, I., & Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology*, *79*(2), 139–147. https://doi.org/10.1016/j.biopsycho.2008.04.001

Jesteadt, W., Bacon, S. P., & Lehman, J. R. (1982). Forward masking as a function of frequency, masker level, and signal delay. *The Journal of the Acoustical Society of America*, *71*(4), 950–962. https://doi.org/10.1121/1.387576

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kelly, A. J., & Dux, P. E. (2011). Different attentional blink tasks reflect distinct information processing limitations: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1867–1873. https://doi.org/10.1037/a0025975

Knight, R. T., & Scabini, D. (1998). Anatomic Bases of Event-Related Potentials and Their Relationship to Novelty Detection in Humans: *Journal of Clinical Neurophysiology*, *15*(1), 3–13. https://doi.org/10.1097/00004691-199801000-00003

Lawrence, M., A. (2016). *ez: Easy Analysis and Visualization of Factorial Experiments*. https://CRAN.R-project.org/package=ez

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). *emmeans: Estimated Marginal Means, aka Least-Square Means* (Version 1.3.1) [R; R (>=3.2)].

https://github.com/rvlenth/emmeans

MacLean, M. H., & Arnell, K. M. (2012). A conceptual and methodological framework for measuring and modulating the attentional blink. *Attention, Perception, & Psychophysics*, *74*(6), 1080–1097. https://doi.org/10.3758/s13414-012-0338-4

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*(6), 657–672. https://doi.org/10.3758/BF03209251

Martens, S., Johnson, A., Bolle, M., & Borst, J. (2009). A Quick Visual Mind Can be a Slow Auditory Mind: Individual Differences in Attentional Selection Across Modalities. *Experimental Psychology*, *56*(1), 33–40. https://doi.org/10.1027/1618-3169.56.1.33

Martens, S., Kandula, M., & Duncan, J. (2010). Restricted Attentional Capacity within but Not between Sensory Modalities: An Individual Differences Approach. *PLoS ONE*, *5*(12), e15280. https://doi.org/10.1371/journal.pone.0015280

Martens, S., Wierda, S. M., Dun, M., de Vries, M., & Smid, H. G. O. M. (2015). Musical Minds: Attentional Blink Reveals Modality-Specific Restrictions. *PLOS ONE*, *10*(2), e0118294. https://doi.org/10.1371/journal.pone.0118294

Meyer, W.-U., Niepel, M., Rudolph, U., & Schützwohl, A. (1991). An experimental analysis of surprise. *Cognition & Emotion*, *5*(4), 295–311. https://doi.org/10.1080/02699939108411042

Mondor, T. A. (1998). A transient processing deficit following selection of an auditory target.

Psychonomic Bulletin & Review, *5*(2), 305–311. https://doi.org/10.3758/BF03212956

Moore, B. C. J., & Glasberg, B. R. (1981). Auditory filter shapes derived in simultaneous and forward masking. *The Journal of the Acoustical Society of America*, *70*(4), 1003–1014. https://doi.org/10.1121/1.386950

Näätänen, R., & Alho, K. (1995). Mismatch negativity-a unique measure of sensory processing in audition. *International Journal of Neuroscience*, *80*(1–4), 317–337. https://doi.org/10.3109/00207459508986107

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Niepel, M., Rudolph, U., Schützwohl, A., & Meyer, W.-U. (1994). Temporal characteristics of the surprise reaction induced by schema-discrepant visual and auditory events. *Cognition & Emotion*, *8*(5), 433–452. https://doi.org/10.1080/02699939408408951

Parmentier, F. B. R. (2014). The cognitive determinants of behavioral distraction by deviant auditory stimuli: A review. *Psychological Research*, *78*(3), 321–338. https://doi.org/10.1007/s00426-013-0534-4

Parmentier, F. B. R., Elsley, J. V., Andrés, P., & Barceló, F. (2011). Why are auditory novels distracting? Contrasting the roles of novelty, violation of expectation and stimulus change. *Cognition*, *119*(3), 374–380. https://doi.org/10.1016/j.cognition.2011.02.001

Pastore, R. E., Harris, L. B., & Goldstein, L. (1980). Auditory forward and backward masking interaction. *Perception & Psychophysics*, *28*(6), 547–549. https://doi.org/10.3758/BF03198823

Pavlov, I. P. (1927). *Conditioned reflexes*. Oxford University Press.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023–1031. https://doi.org/10.3758/s13428-013-0434-y

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pinto, Y., van der Leij, A. R., Sligte, I. G., Lamme, V. A. F., & Scholte, H. S. (2013). Bottom-up and top-down attention are independent. *Journal of Vision*, *13*(3), 16–16. https://doi.org/10.1167/13.3.16

Potter, M. C., & Chun, M. M. (1998). *Two Attentional Deficits in Serial Target Search: The Visual Attentional Blink and an Amodal Task-Switch Deficit*. 14.

Qualtrics. (2005). *Qualtrics*. Qualtrics. http://www.qualtrics.com

Rauschenberger, R. (2010). Reentrant processing in attentional guidance—Time to abandon old dichotomies. *Acta Psychologica*, *135*(2), 109–111. https://doi.org/10.1016/j.actpsy.2010.04.014

Roeber, U., Berti, S., & Schröger, E. (2003). Auditory distraction with different presentation rates: An event-related potential and behavioral study. *Clinical Neurophysiology*, *114*(2), 341–349. https://doi.org/10.1016/S1388-2457(02)00377-2

Ruthruff, E., & Pashler, H. (2001). Perceptual and central interference in dual-task performance. In K. Shapiro (Ed.), *The limits of attention: Temporal constraints in human information processing* (pp. 100–123). Oxford University Press.

Salmi, J., Rinne, T., Koistinen, S., Salonen, O., & Alho, K. (2009). Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention. *Brain Research*, *1286*, 155–164. https://doi.org/10.1016/j.brainres.2009.06.083

Scharf, B. (1998). Auditory attention: The psychoacoustical approach. In H. Pashler (Ed.), *Attention* (pp. 75–117). Psychology Press.

Schröger, E. (1996). A Neural Mechanism for Involuntary Attention Shifts to Changes in Auditory Stimulation. *Journal of Cognitive Neuroscience*, *8*(6), 527–539. https://doi.org/10.1162/jocn.1996.8.6.527

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychological Science*, *16*(2), 114–122.

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to Study Clinical Populations. *Clinical Psychological Science*, *1*(2), 213–220. https://doi.org/10.1177/2167702612469015

Shen, D., & Alain, C. (2012). Implicit Temporal Expectation Attenuates Auditory Attentional Blink. *PLoS ONE*, *7*(4), e36031. https://doi.org/10.1371/journal.pone.0036031

Shen, D., & Mondor, T. A. (2006). Effect of distractor sounds on the auditory attentional blink. *Perception & Psychophysics*, *68*(2), 228–243. https://doi.org/10.3758/BF03193672

Shen, D., & Mondor, T. A. (2008). Object file continuity and the auditory attentional blink. *Perception & Psychophysics*, *70*(5), 896–915. https://doi.org/10.3758/PP.70.5.896

Shore, D. I., McLaughlin, E. N., & Klein, R. M. (2001). Modulation of the attentional blink by differential resource allocation. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *55*(4), 318–324. https://doi.org/10.1037/h0087379

Sokolov, E. N. (1990). The orienting response, and future directions of its development. *The Pavlovian Journal of Biological Science*, *25*(3), 142–150. https://doi.org/10.1007/BF02974268

Sokolov, E. N. (Ed.). (2002). *The orienting response in information processing*. L. Erlbaum Associates.

Soto-Faraco, S., & Spence, C. (2002). Modality-specific auditory and visual temporal processing deficits. *The Quarterly Journal of Experimental Psychology Section A*, *55*(1), 23–40. https://doi.org/10.1080/02724980143000136

Spearman, C. (2010). The proof and measurement of association between two things.

International Journal of Epidemiology, *39*(5), 1137–1150. https://doi.org/10.1093/ije/dyq191

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Summerville, A., & Chartier, C. R. (2013). Pseudo-dyadic “interaction” on Amazon’s Mechanical Turk. *Behavior Research Methods*, *45*(1), 116–124. https://doi.org/10.3758/s13428-012-0250-9

Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin & Review*, *10*(3), 630–637. https://doi.org/10.3758/BF03196525

Theeuwes, J., Atchley, P., & Kramer, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. In S. Monsell & J. Driver (Eds.), *Control of Cognitive Processes: Attention and Performance Xviii* (ISI:000175879400004; pp. 105–124). MIT Press. ://000175879400004

Tremblay, S., Vachon, F., & Jones, D. M. (2005). Attentional and perceptual sources of the auditory attentional blink. *Perception & Psychophysics*, *67*(2), 195–208.

Vachon, F., Hughes, R. W., & Jones, D. M. (2012). Broken expectations: Violation of expectancies, not novelty, captures auditory attention. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *38*(1), 164–177. https://doi.org/10.1037/a0025054

Vachon, F., Labonté, K., & Marsh, J. E. (2017). Attentional capture by deviant sounds: A noncontingent form of auditory distraction? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 622–634. https://doi.org/10.1037/xlm0000330

Vachon, F., & Tremblay, S. (2005). Auditory Attentional Blink: Masking the Second Target is Necessary, Delayed Masking is Sufficient. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *59*(4), 279–286. https://doi.org/10.1037/h0087480

Vachon, F., & Tremblay, S. (2006). Delayed Masking and the Auditory Attentional Blink: A Test for Retrieval Competition and Bottleneck Models. *Experimental Psychology*, *53*(3), 182–190. https://doi.org/10.1027/1618-3169.53.3.182

Vachon, F., Tremblay, S., Hughes, R. W., & Jones, D. M. (2009). Capturing and Unmasking the Mask in the Auditory Attentional Blink. *Experimental Psychology*, *57*(5), 346–353. https://doi.org/10.1027/1618-3169/a000041