


Cat-astrophic effects of sudden interruptions on spatial auditory attention

Wusheng Liang,¹  Christopher A. Brown,²  and Barbara G. Shinn-Cunningham^{3,a)} 

¹Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

²Department of Communication Science and Disorders, The University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA

³Neuroscience Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

ABSTRACT:

Salient interruptions draw attention involuntarily. Here, we explored whether this effect depends on the spatial and temporal relationships between a target stream and interrupter. In a series of online experiments, listeners focused spatial attention on a target stream of spoken syllables in the presence of an otherwise identical distractor stream from the opposite hemifield. On some random trials, an interrupter (a cat “MEOW”) occurred. Experiment 1 established that the interrupter, which occurred randomly in 25% of the trials in the hemifield opposite the target, degraded target recall. Moreover, a majority of participants exhibited this degradation for the first target syllable, which finished before the interrupter began. Experiment 2 showed that the effect of an interrupter was similar whether it occurred in the opposite or the same hemifield as the target. Experiment 3 found that the interrupter degraded performance slightly if it occurred before the target stream began but had no effect if it began after the target stream ended. Experiment 4 showed decreased interruption effects when the interruption frequency increased (50% of the trials). These results demonstrate that a salient interrupter disrupts recall of a target stream, regardless of its direction, especially if it occurs during a target stream. © 2022 Acoustical Society of America.

<https://doi.org/10.1121/10.0010453>

(Received 3 March 2022; revised 22 April 2022; accepted 25 April 2022; published online 16 May 2022)

[Editor: Matthew J. Goupell]

Pages: 3219–3233

I. INTRODUCTION

Solving the cocktail party problem (Cherry, 1953), refers to the ability to focus attention on one sound source amidst temporally overlapping, competing sounds (Fritz *et al.*, 2007; Shinn-Cunningham *et al.*, 2017; Sussman, 2017). However, attention in any sensory modality is a “biased competition” (Desimone and Duncan, 1995), a fight between focusing attention volitionally on whatever source seems most relevant and the involuntary reorientation of attention towards some inherently salient input. Even though attention in noisy, social settings depends strongly on disruptions from this latter, bottom-up process, we know relatively little about how salient unattended events affect perception (although see, for instance, Kaya and Elhilali, 2014).

Salient inputs in any sensory modality can draw attention involuntarily, whether it is from a flash of lightning, the honk of a car in your blind spot, or the feel of a spider crawling onto your hand. Salience itself depends on the recent history of inputs: events that are different from others in a scene and that therefore stand out as new or unexpected will draw attention involuntarily. For instance, if you found yourself in traffic gridlocked due to revelers celebrating after your home team’s World Series win, a honking car likely fits in and would not draw attention strongly. On the

other hand, in simpler, controlled auditory scenes, even a change in the loudness of a sound can grab attention (Salmi *et al.*, 2009). A sound perceived as an entirely new object, like the sudden ringing of a phone, almost always reorients attention involuntarily (Kaya and Elhilali, 2014). Such disruptions interfere with analysis and recall of an object that a listener may be trying to listen to; for instance, they impair target detection accuracy (Salmi *et al.*, 2009).

These examples are consistent with the idea that perceptual objects, the brain’s estimate of the sound energy originating from a single physical sound source, serve as the basic perceptual unit of auditory attention (Shinn-Cunningham, 2008). Both top-down and bottom-up attention depends on how the brain organizes sound, perceptually segregating the acoustic mixture into different perceptual objects (Bregman, 1994). When we focus on a particular auditory object, whether voluntarily or involuntarily, other competing sounds are relegated to the perceptual background (Duncan, 2006; Shinn-Cunningham, 2008). Typically, although listeners may be aware of the presence of unattended objects in a scene, they cannot easily recall the contents or features of objects that are not in the attentional foreground (Goldstein and Fink, 1981; Neisser and Becklen, 1975; Rock and Gutman, 1981). On the other hand, even features of an attended object that are unimportant for a task are likely to be perceptible. For instance, listeners are typically aware of the timbre of the voice speaking the words that they have been asked to repeat back

^{a)}Electronic mail: bgsc@andrew.cmu.edu

or the orientation of a line that they were asked to locate in a visual scene. Moreover, making multiple judgments of features within the same object is easy, while making such judgments across different objects is not (Best *et al.*, 2008; Duncan, 1984; Marinato and Baldauf, 2019); see review by Chen (2012). When new objects appear, they involuntarily draw attention.

To understand the influences of bottom-up attention on auditory performance, many experiments ask listeners to focus volitional, top-down attention on a target source and then measure the impact of presenting a salient but task-irrelevant sound during the target. Despite the fact that spatial location provides a strong cue for guiding top-down, volitional attention (e.g., Arbogast and Kidd, 2000; Broadbent, 1954; Mondor and Zatorre, 1995), most past studies of bottom-up auditory attention presented auditory scenes without differences in the locations of the competing sources, using either monaural (Salmi *et al.*, 2009) or diotic (Huang and Elhilali, 2020) headphone presentations. Few if any have explored whether bottom-up effects of an interrupting sound depend on the spatial configuration of the target and interrupting sources.

How the spatial configuration of an interrupting sound might disrupt top-down spatial attention is unclear. One might postulate that interrupters that are spatially near the target source are less disruptive: when listeners must shift volitional spatial attention from one source to a different source, the time required to refocus attention increases linearly with the angular separation of the sources (Mondor and Zatorre, 1995; Rhodes, 1987). Given this, if an interruption comes from a direction similar to the location of the target, it might take less time to reorient attention to the interruption (and return attention to the target) compared to an interrupter that is spatially more distant from the target, leading to less interference. On the other hand, in a multi-talker environment, the degree to which spatially focused attention suppresses a background sound increases with the spatial separation between the target and distractor (Allen *et al.*, 2009; Best *et al.*, 2006; Brown, 2014). This suggests that interruptions coming from a direction similar to the target will be suppressed less effectively and therefore interfere more with understanding the target compared to interrupters from further away, which may be more fully suppressed. Finally, it may be that bottom-up interruptions of salient events operate independently of the direction of top-down spatial attention, depending instead only on the salience of the interrupter. Consistent with this, in visual studies, an abrupt interrupter interferes with top-down attention regardless of task goals (e.g., see Folk and Remington, 2015). Thus, there are three opposing hypotheses for how the spatial location of the interrupter relative to the target might affect target recall:

(1) interrupters contralateral to the target should be more disruptive than ipsilateral interrupters if the disruption depends on the time it takes to reorient attention across space;

(2) interrupters ipsilateral to the target should be more disruptive than contralateral interrupters if top-down spatial attention is responsible for suppressing the interrupter; and

(3) contralateral and ipsilateral interrupters should be equally disruptive if their effects depend only on their salience, and not on top-down attention.

Bottom-up disruptions of attention are thought to be driven by a violation of expectation signaled by an abrupt perceptual change (Parmentier *et al.*, 2011). The sudden onset of a distinct new object within a trial is a form of such violation that can grab attention involuntarily. However, visual studies have shown that abrupt interrupters cause less interference if they are encountered on a higher percentage of trials compared to when they occur only rarely (e.g., see Müller *et al.*, 2009). Thus, at least in vision, longer-term expectations influence how disruptive an interrupter will be; specifically, disruption does not depend only upon whether an event is distinct and new within the context of a trial, but upon whether the event is expected or unexpected across trials. Little is known about whether high-level expectations modulate how much a sudden auditory event disrupts recall of a target auditory stream.

To investigate how the location and timing of an abrupt interrupter disrupts top-down spatial attention to and recall of an auditory stream, we conducted four online studies. In each case, the target was a sequence of speech syllables spoken by a male talker, while the interrupter was quite distinct: a cat “meow.” With this design, the interrupter was always perceived as a new auditory object within each trial on which it occurred and thus was likely to be a salient, distinct event. The first experiment established that when the interrupter was played after the target began and before it was completed, target recall was disrupted if the interrupter was unexpected (occurring on only 25% of trials) and contralateral to the target. The second experiment found that unexpected ipsilateral and contralateral interrupters occurring during the target stream were equally disruptive to target recall. The third showed that an unexpected interrupter weakly disrupted target recall when it occurred before the first target syllable, but had no effect if it came just after the final target syllable. Finally, the fourth demonstrated that an expected interrupter (occurring in half of all trials) caused less disruption of target recall than the unexpected interrupter used in the first three experiments.

II. METHODS

A. Overview

Four experiments shared the same basic task. In all experiments, participants heard two competing sequences of syllables, one from each side of midline (Fig. 1): a target speech stream to which they were asked to attend, and a competing speech stream (the “distractor”) from the opposite hemifield. The syllables in the target and distractor streams were from the same male speaker; thus, all of the

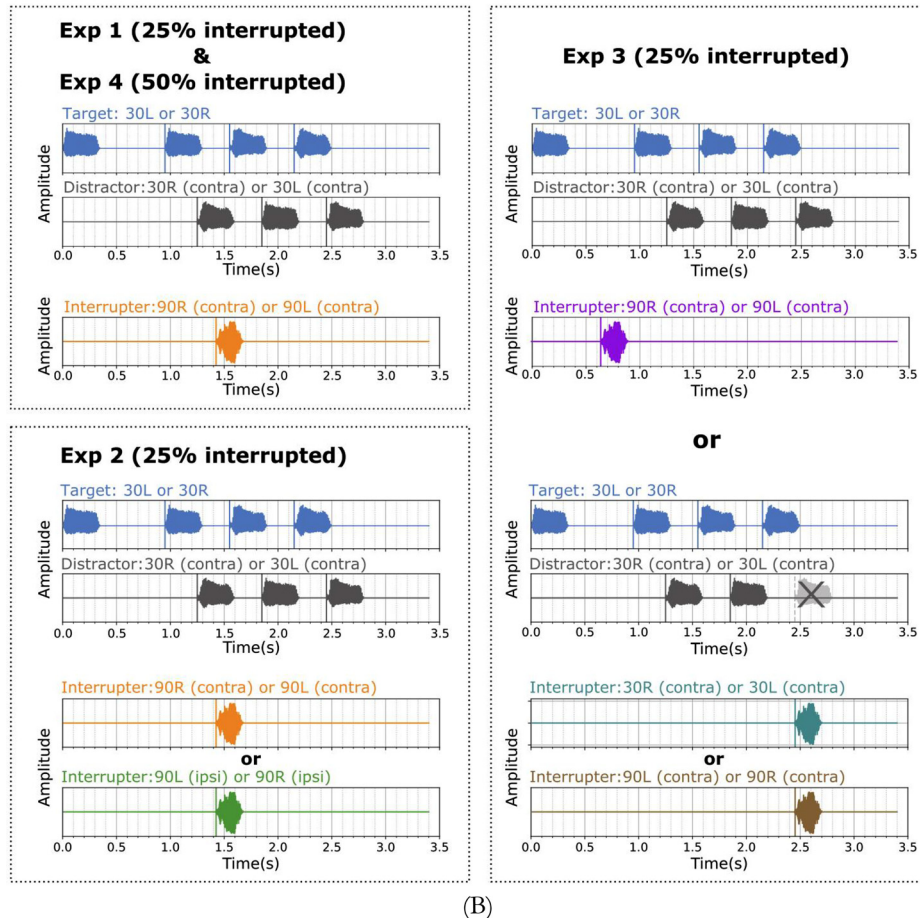
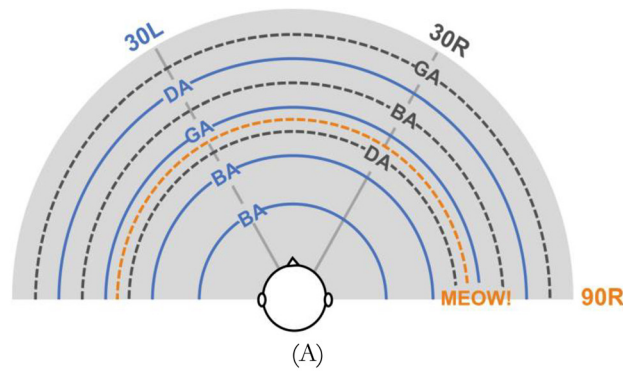


FIG. 1. (Color online) (A) illustrates the spatial layout and rough timing, while (B) shows the timing of each stream in more detail. (A) Schematic of the stimuli for example trials that could occur in Experiments 1, 2, and 4. An auditory cue was first presented from the target direction, then a 3-syllable target stream and a 3-syllable distractor stream were presented from 30° azimuth in opposite hemifields (in the example, the target is to the left and the distractor to the right). Some percentage of random trials contained interrupters, a salient and transient cat MEOW. The percentage of trials containing MEOWS, as well as the timing and direction of the MEOW depended on the experiment [see (B)]. However, the example here, where the MEOW comes after the first and before the second target syllable and from 90° contralateral (contra) to the target, was the most common condition and occurred in Experiments 1, 2, and 4. (B) Timing diagrams for the different trial types in each experiment. Experiments 1 and 4 always presented the interrupter from 90° contralateral (contra) to the target stream, after the first and before the second target syllable, but the two experiments differed in the likelihood that an interrupter occurred (25% and 50%, respectively). Experiment 2 was like Experiment 1, except that the interrupter was equally likely to be 90° contralateral or 90° ipsilateral to the target. Experiment 3 had three equally likely interrupter conditions, which together were 25% of all trials. An early interrupter occurred after the cue and before the first target syllable, from 90° contralateral to the target. When late interrupters occurred, they replaced the final distractor syllable, and either were at 30° or 90° contralateral to the target.

experiments required top-down spatial attention toward the target in order for listeners to avoid confusion with the distractor. At the start of each trial, an auditory cue indicated the direction of the sequence that was the target. At the end of the trial, listeners were asked to report the target syllables coming from the attended, target direction. On a subset of trials, an interrupting sound (a cat “meow”) was presented. The experiments differed in the exact spatial configuration and timing of this “interrupter.”

In Experiment 1, the interrupter was always presented before the 2nd target syllable, and from the side contralateral to the attended target stream. We hypothesized that in this experiment, the interrupter would interfere with the identification of the target syllables, especially the second

syllable that was presented just after the interrupter. Experiment 2 was identical to Experiment 1 except that the interrupter was randomly and equally likely to be located either contralateral to the target (as in Experiment 1) or ipsilateral. Experiment 2 examined whether performance depended on the spatial configuration of the interrupter relative to the attended target. In Experiment 3, the interrupter was presented from the same hemifield as the distractor, but it was presented either prior to the first syllable or after the final syllable in the target stream, to explore whether interference in recall of target syllables was greater when the interrupter occurs in the middle of the target stream. This experiment was designed specifically to determine whether the backwards-in-time interference we saw in the first two

experiments was eliminated if the target stream completed playing before the interrupter occurred. Experiment 4 was identical to Experiment 1 except that the interrupter likelihood increased to 50% on each trial. This final experiment allowed us to see if the interrupter interfered when it was highly likely to occur; if so, it would allow planned future experiments on the underlying neural bases of interference effects to be run more efficiently, with fewer total interrupted trials, as it would take less time to gather a sufficient number of interrupter trials to obtain neural measures.

B. Participants

In Experiment 1, data were collected from 45 participants between the ages of 19 and 64 [mean = 31.0, standard deviation (std) = 10.1, 24 females, 21 males]. In Experiment 2, data were collected from 45 participants between the ages of 18 and 56 (mean = 31.7, std = 9.8, 21 females, 24 males). In Experiment 3, data were collected from 40 participants between the ages of 20 and 64 (mean = 35.2, std = 11.8, 14 females, 24 males, 2 not provided). In Experiment 4, data were collected from 45 participants between the ages of 18 and 65 (mean = 31.3, std = 10.0, 21 females, 21 males, 3 not provided). No participants completed more than one of the experiments. All of the participants, recruited through the online Prolific portal, were native English speakers with self-reported normal hearing. Experiments were run online using the Gorilla platform after being approved by the Carnegie Mellon University Institutional Review Board. All participants provided informed consent before participating in the study and were paid for their participation.

C. Stimuli

Target and distractor streams each consisted of three consonant-vowel syllables (/ba/, /da/, and /ga/), presented in randomly permuted order. We chose to construct the target stream from speech syllables to avoid having to train listeners to identify arbitrary sounds. We used a small, closed set of syllables in the identification task to mitigate demands on working memory capacity since our interest was in exploring how interrupters interfere with top-down attention and with transferring perceived sounds into working memory. We selected this particular set of plosive syllables because their abrupt onsets have previously been shown to evoke clean event-related potentials whose magnitudes are modulated by selective attention (Deng *et al.*, 2019); in the near future, we intend to study the neural effects of bottom-up interruptions on these evoked responses using a similar behavioral paradigm while measuring electroencephalography.

The same syllables were used for Experiments 1, 2, and 3 (each 0.45 s in duration). At the conclusion of Experiment 3, we determined that some participants confused some syllables that were not sufficiently distinct, acoustically, in the original set. Given that these confusions affected all conditions equally, they did not alter our conclusions; however, for the final experiment, we re-recorded the syllables to make them more easily distinguishable. The syllables used

in Experiment 4 were each 0.35 s in duration. Each set of syllables was recorded by a single male talker whose native language was English (although the talkers differed in the two sets). Both talkers had similar fundamental frequencies, near 100 Hz.

The interrupter was a cat MEOW sound, which was only present in some randomly chosen trials (25% in Experiments 1, 2, and 3; 50% in Experiment 4). The MEOW (0.25 s) was retrieved from the internet.

D. Spatialization and signal levels

Both the syllable streams and the MEOW interrupter were spatialized using pseudo-anechoic, non-individualized head-related impulse responses (HRIRs). HRIRs were derived from recordings from the ears of a female volunteer in a medium-sized office at Carnegie Mellon University. The RT60 of the room was measured using Schroeder's method (Schroeder, 1965) to be 0.6 s or less over the frequency range of interest. Specifically, we first measured room-related impulse responses (RRIRs) that contained natural reverberant energy, then time windowed the measurements to find the HRIRs. When measuring the RRIRs, the loudspeaker (MSP5A, Yamaha, Shizuoka, Japan) was placed 1.5 m away from the subject's head at 0° elevation. The individual sat about 1.5 m away from walls or other hard surfaces, and a carpet covered the floor between the loudspeaker and the listener to attenuate the echo off the floor. A pair of KE-4-211-2 4.75 mm electret condenser microphone capsules (Sennheiser, Hannover, Germany) were mounted into hollowed-out foam earplugs and positioned at the entrance of the ear canals of the individual. A 5-s sweep signal (50 Hz–18 kHz) played from the loudspeaker; the response was measured at the ear canal entrance.

RRIRs were measured for azimuths of −90°, −30°, 30°, and 90° relative to the listener; these were obtained by rotating the chair, keeping the other equipment the same. Microphone outputs were matched by subtracting the level difference measured from a reference recording, which was obtained by placing the microphones in front of the loudspeaker in close proximity to each other and recording a broadband reference sound. RRIRs were recovered by convolving the recorded signals with the inverse filter of the sweep signal. To isolate the direct sound impulse response and remove echoes and reverberation, the RRIRs were time windowed to 4.5 ms duration by multiplying with a time window with 0.4 ms-long Hanning onset ramp and a 1 ms-long Hanning offset ramp that started 0.5 ms before the RRIR peak. Based on visual inspection, this window captured all of the direct-sound impulse response that was above the noise floor and windowed out reflected energy. Spatialized stimuli were generated by convolving HRIRs for the desired azimuthal locations with the original syllables and the cat sound.

In Experiments 1 and 2, the spatialized syllables were scaled to have a peak value of 0.5 in the ipsilateral ear; as a result, the /ba/ was slightly more intense than the other two syllables (by 2.9 dB re: /da/; by 2.1 dB re: /ga/) across all of

the tested conditions. While this difference might have made /ba/ slightly more understandable, this effect is consistent across all conditions in the affected experiments and thus should have no effect on the comparisons of interest. To make the interrupter more salient, the MEOW was presented at a level 7.4 dB more intense than the /ba/ (i.e., 10.3 dB more than /da/, and 9.5 dB more than /ga/).

In Experiments 3 and 4, the source syllables were first adjusted to be the same root-mean-square (rms) level and then convolved with the HRIRs, which were also controlled to have the same rms level on the louder channel across different simulated directions. To make the interrupter more salient, the MEOW level was set to be 8 dB more intense than the syllables before being convolved by the appropriate HRIR.

In all four experiments, the final spatialized stimuli were created by summing the target stream, distractor stream, and interrupter at the appropriate times (see Sec. II E). The final signals were then all attenuated by the same amount so that the peak magnitude across the entire set of stimuli was less than one to avoid clipping. Ultimately, online participants were asked to set the sound levels of the stimuli to be at a comfortable presentation level at the start of each experiment, so the absolute levels of the stimuli were not controlled beyond this.

E. Main task

In each of the experiments, participants were instructed to listen at the start of each trial for an auditory cue, which was a single /ba/ syllable, spatialized to come from either -30° (to the left) or $+30^\circ$ (to the right). Participants were asked to report back the sequence of three target syllables played subsequently from the cued direction (the target stream). The cue and target directions were randomly and independently selected on each trial. A competing 3-syllable distractor stream was played from the hemifield opposite the cue and target stream (either $+30^\circ$ or -30° , respectively), except for trials in Experiment 3, where late interrupters were presented *instead of* the final distractor syllable so that the distractor stream consisted of only the initial two syllables.

On each trial, the target and distractor streams were each made up of a random sequence of syllables selected from the same set of /ba/, /da/, and /ga/ syllables, without replacement (similar to Deng *et al.*, 2019), yielding six possible permutations of the three syllables in the target stream and (separately) in the distractor stream. The onsets of the syllables within each stream were separated by 600 ms. The target and distractor were temporally interleaved: the target stream started 500 ms after completion of the spatial cue, while the first distractor syllable began 800 ms after the completion of the spatial cue (300 ms after the first target syllable).

All trials contained both target and distractor streams from opposite hemifields. However, in addition, some percentage of randomly selected trials also contained a 0.25 s-long “interrupter” (MEOW). In Experiments 1, 2, and 3, 25% of the trials contained an interrupter, while in Experiment 4, 50% of trials contained an interrupter (cat MEOW).

In Experiments 1, 2, and 4, the interrupter always began 125 ms before the onset of the second target syllable (475 ms after the start and about 25 ms after completion of the first target syllable). In these experiments, the interrupter was spatialized to either -90° or 90° in azimuth, depending on the trial and experiment. In Experiments 1 and 4, the interrupter was always presented from the same hemifield as the distractor, but from a lateral angle of 90° . In Experiment 2, the interrupter had a lateral angle of 90° but was equally likely to come from the hemifield ipsilateral to the distractor and the hemifield contralateral to the distractor, randomly selected on each trial where it occurred.

In Experiment 3, we tested both “early” and “late” interrupters, neither of which overlapped temporally with the target stream. Specifically, the interrupter either (1) was presented 300 ms before the onset of the 1st target syllable from 90° contralateral to the target stream, (2) temporally replaced the 3rd distractor syllable (starting 300 ms after the onset of the 3rd target syllable) from 90° contralateral to the target stream (farther to the side than the distractor syllables), or (3) temporally replaced the 3rd distractor syllable from 30° contralateral to the target stream (i.e., at the same location as the distractor syllables). Thus, in Experiment 3, the late interrupters were closer in time to the onset of the final target syllable (starting 300 ms after the final target syllable began) than the interrupter was to the onset of the first syllable in the other experiments (475 ms), which showed “backwards in time” interference. Figure 1 shows a schematic for a trial in which the interrupter occurs 125 ms before the onset of the 2nd target syllable from 90° contralateral to the target direction, consistent with interrupter trials that could occur in Experiments 1, 2, and 4.

Following the presentation of the sounds, participants were asked to report the target syllable sequence by clicking on buttons on a graphical user interface (GUI). Responses were not constrained to include each syllable; instead, participants were allowed to respond with the same syllable multiple times, for different serial positions, on any given trial. No feedback was provided during the task session. The next trial began automatically 0.5 s after the participants pressed a “continue” button after entering their response to the current trial.

F. Experimental procedure

Because participants were required to perform the spatial auditory attention task over the internet using their own headphones, a headphone screening was conducted at the start of each session using a Huggins pitch stimulus (Milne *et al.*, 2021). This test ensured that participants used headphones and a playback system that preserved the interaural time differences between the ears. During the headphone check, the participants were asked to put on their headphones and listen to three stereo white noise stimuli for each trial, two of which were diotic and one of which was diotic except for a narrowband noise with an interaural phase shift of 180° , leading to a percept of pitch in that interval if and

only if their playback system preserved interaural cues. The participants then were asked to identify which of the intervals contained the Huggins pitch. Participants who failed to respond correctly for all of the six trials within three tries were rejected.

Participants who successfully passed the Huggins pitch screening then were allowed to continue to the main experiment. The main experiment began with a 6-trial training session in which participants used the GUI to report the target syllables from uninterrupted trials; only participants who correctly reported all three of the target syllables in at least 4 of the 6 trials could proceed to the main study. Feedback was provided after each trial of the training session with the correct target syllables shown on the screen.

For all experiments, the target stream was equally likely to come from the left or the right on each trial. The direction was pseudo-random and independent from trial to trial, other than the constraint that left and right trials were presented an equal number of times in each experimental block.

For Experiment 1, each subject completed 200 randomized trials, 100 of which had the target to the left and 100 with the target to the right. Twenty-five of the 100 trials on each side were interrupted trials. The 200 trials were organized into four blocks, each containing 50 trials. At the end of each block, a screen prompted participants to take a break if they desired.

For Experiment 2, each subject performed 192 randomized trials in total. On each interrupted trial, the interrupter was randomly chosen, with equal likelihood, to be either ipsilateral or contralateral to the distractor (i.e., 48 trials had interrupters, of which 24 had an interrupter on the same side as the distractor and 24 had an interrupter on the opposite side). The 192 trials were organized into four blocks of 48 trials, with breaks offered at the end of each block.

For Experiment 3, each subject performed 288 trials in total (72 interrupted trials and 216 uninterrupted trials); in the 72 interrupted trials, the interrupter was equally likely to be presented early from 90°, late from 90°, or late from 30° contralateral to the target direction (i.e., there were 72 interrupted trials comprising 24 trials for each of the three interrupter conditions). The 288 trials were organized into six blocks of 48 trials, with the screen indicating that participants could take a break at the end of each block.

Experiment 4 was identical to Experiment 1 (interrupter from 90° contralateral to the target and occurring after the first and before the second target syllable), except that 50% of the trials contained an interrupter. Each subject completed 96 randomized trials, of which 48 had interrupters. The 96 trials were organized into two blocks of 48 trials, separated by break prompts.

G. Data analysis

In each experiment, behavioral performance was quantified by first computing the raw percentage of trials in which each syllable in the target stream was correctly identified. We anticipated that performance without an interrupter would

likely vary with syllable position. For instance, when recalling a list of items, both the initial and the final items tend to be remembered more accurately than interior items (termed primacy and recency effects, respectively), regardless of other manipulations (e.g., see [Murdock, 1962](#)). The distractor stream may also cause interference, both by distracting attention away from the target when it begins (before the second syllable) and by causing energetic masking of the target syllables (e.g., see [Kidd et al., 2005](#)). However, our main interest was in the effect of the occasional, distinct MEOW interrupter. Thus, we treated the uninterrupted condition in each experiment as a within-subject baseline for recall of each of the three target syllables. To summarize the effects of the interrupter, for each subject we computed the effects of each type of interrupter on performance separately for each target syllable by subtracting the raw percent correct in the interrupted trials from the percent correct in the uninterrupted trials. In all experiments and conditions, performance differences were deemed sufficiently Gaussian to be analyzed using t-tests and analysis of variance (ANOVA). Specifically, for each data group in the performance difference data (grouped by syllable position and interrupter type), Shapiro-Wilk tests were performed and histograms were plotted to confirm the normality of the data.

We were mainly interested in two questions: (1) for which syllables did an interrupter degrade accuracy of recall? (2) If there were multiple conditions in which interrupters had significant effects, was the size of this effect significantly different across conditions? To answer the first question, we conducted one-tailed t-tests on the performance-difference data to determine whether the interrupter led to worse performance than when there was no interrupter (i.e., we expected positive performance differences). We did this separately for each type of interrupter and each syllable, with Bonferroni correction for multiple comparisons. To test whether any significant effects of the interrupter identified by our t-tests varied with syllable position and interrupter type, we ran subsequent single or multi-way ANOVAs (depending on the number of independent parameters in the experiment) on the performance difference data as needed (i.e., in cases when there were multiple syllables/conditions where the interrupter had a significant effect). *Post hoc* Tukey honestly significant difference (HSD) tests were conducted to interpret significant effects in these ANOVAs. Thus, for Experiment 1, a repeated-measures one-way ANOVA was conducted on the performance differences with the main factor of syllable position (1st, 2nd, or 3rd). For Experiment 2, a repeated-measures 2-way ANOVA was conducted on the performance differences with the main factors of syllable position (1st, 2nd, or 3rd) and interrupter direction (contralateral to the target or ipsilateral to the target). For Experiments 3 and 4, no follow-up ANOVA was necessary (see Sec. III).

III. RESULTS

A. Experiment 1

Experiment 1 compared participants' raw percent correct performance with and without the interrupter, which when present came from the hemifield contralateral to the

target syllable stream, separately for each syllable (Fig. 2). Overall, the interrupter impaired target syllable identification performance, with the greatest effect occurring on the second syllable [in Fig. 2(A), performance for each of the syllables was worse on interrupted trials, in orange, than on uninterrupted trials, in blue]. This pattern was consistent across individual participants. Figure 2(B) shows this by plotting, for each of the target syllables, individual results connected by lines that are colored according to which condition, uninterrupted or interrupted, led to better performance. In Fig. 2(B), the vast majority of participants performed better on uninterrupted than in interrupted trials (blue lines outnumber orange in all plots, especially for the second and third target syllables). Specifically, performance was better in 43 of 45 participants for syllable 2 and 38 of

45 for syllable 3; even for the first syllable, 34 of 45 participants performed better without the interrupter. Figure 2(C) shows the decrease in percent correct due to the contralateral interrupter for the three syllables. The greatest effect occurred on the second syllable. This is not surprising, given that the second syllable is temporally closer to the interrupter than any other syllable. In addition, the second syllable is the only one that temporally overlaps with the interrupter (approximately 125 ms at the end of the meow temporally overlaps with start of the second target syllable). Both of these factors could lead to a larger disruption of recall for the second target syllable. Figure 2(D) plots these differences for the individual participants for each syllable. In Fig. 2(D), the line segment type (solid or dashed) denotes whether the effect of the interrupter on syllable 2 was larger

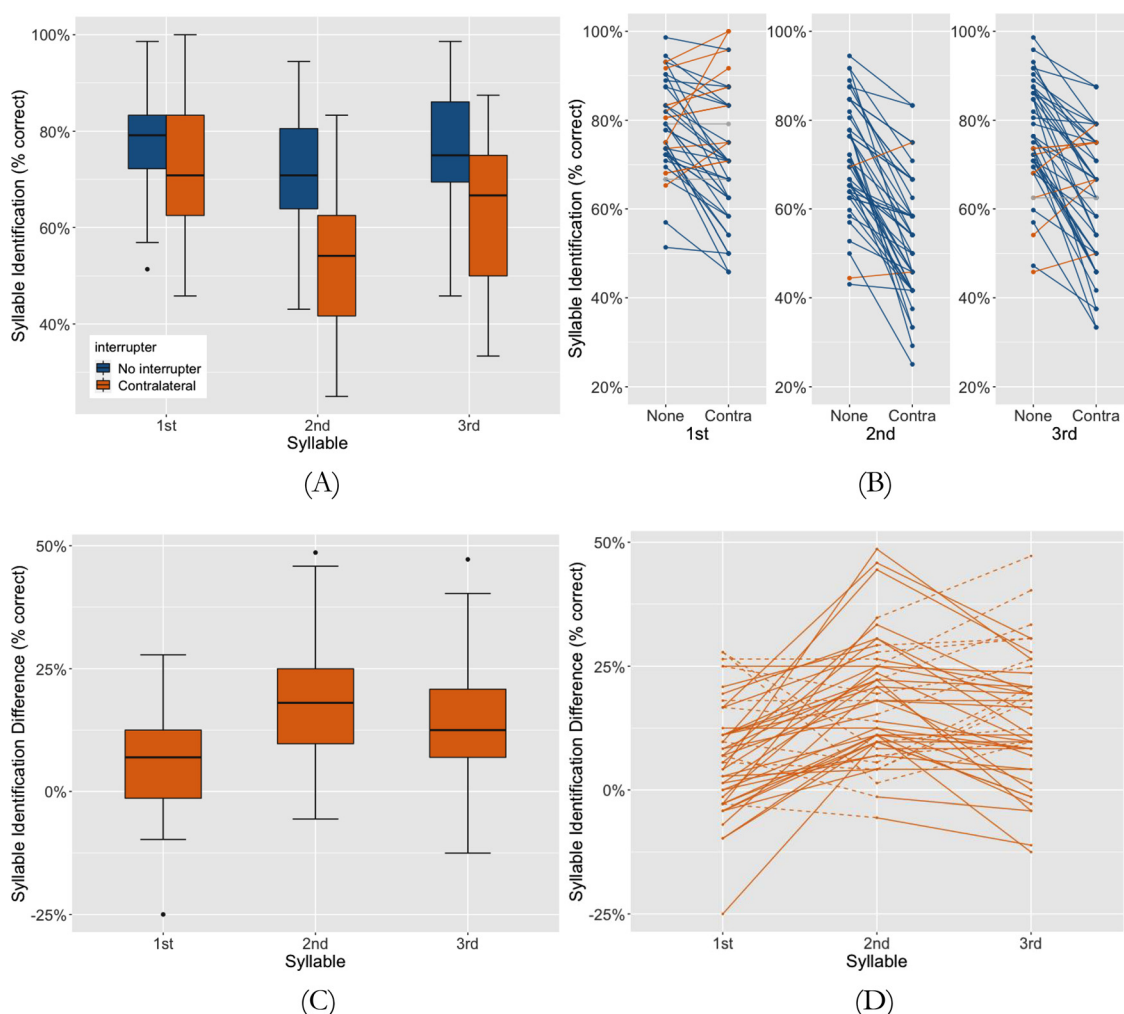


FIG. 2. (Color online) Results for Experiment 1 ($N=45$). (A) Group-level raw percent correct recall for each of the syllables when they are uninterrupted (blue) and interrupted (orange). Box plots cover from 25th to 75th percentile of performance, while range is shown by whiskers. Outliers are shown by asterisks. (B) Raw percent correct recall plotted for individual participants for each of the syllables with and without the interrupter. The line color indicates whether, for that syllable, performance for a given subject is better for the uninterrupted condition (blue) or for the interrupted condition (orange). (C) Group-level within-subject difference in percent recall, showing how large a decrement in performance the interrupter caused for each syllable relative to performance in the uninterrupted condition (baseline), with boxplot conventions as in (A). T-tests on these group level data reveal a significant effect of the interrupter on each syllable. (D) Difference in percent recall re: uninterrupted condition plotted for individual participants. Lines connect individual differences for each syllable. The format of each line segment indicates whether for the pair of connected syllables, the effect of the interrupter was greater for syllable 2 (solid) or syllables 1 or 3 (dashed). Follow-up statistical tests confirm that the effect of the interrupter varies with syllable position, and is significantly smaller for syllable 1 than for syllables 2 or 3.

or smaller (respectively) than its effect on syllable 1 (left line segments) and on syllable 3 (right line segments). This plot shows that the majority of participants showed larger effects of the interrupter on recall of syllable 2 than of syllable 1 (35 of 45 participants) and for recall of syllable 2 than of syllable 3 (31 of 45 participants).

Statistical analyses confirmed these observations. Performance in the interrupted trials was significantly worse than in uninterrupted trials on the 2nd ($t_{44} = 9.786, p < 0.001$), 3rd ($t_{44} = 7.254, p < 0.001$), and even the 1st syllable ($t_{44} = 4.048, p < 0.001$). Subsequent repeated-measures one-way ANOVA on the difference data supported the conclusion that the influence of the interrupter varied with syllable position ($F_{2,88} = 16.871, p < 0.001$); *Post hoc* Tukey tests showed that the interruption effect was significantly bigger on the 2nd ($p < 0.001$) and 3rd ($p < 0.001$) syllables compared with the 1st syllable, but not significantly different for the second and third syllables.

B. Experiment 2

Experiment 2 was designed to determine whether the effect of the interrupter depended on its direction relative to the direction of the target stream. To address this question, in half of the interrupted trials the interrupter occurred at 90° from the midsagittal plane in the same hemifield as the target, and in the other half, it was presented at 90° in the other hemifield (Fig. 3).

Figure 3(A) shows that, as in Experiment 1, performance tended to be worse for the second syllable than either of the other syllables and also worse when the interrupter was present. Importantly, there was no noticeable effect of the interrupter location. Figure 3(B) further confirms that the majority of participants performed better in the uninterrupted trials than when there was an interrupter on the contralateral hemifield or the ipsilateral hemifield (the majority of the line segments are blue, showing that performance was better in the uninterrupted condition than the two interrupted conditions). Specifically, out of 45 participants, the number who performed worse with an ipsilateral interrupter than no interrupter was 31 for the first syllable, 40 for the second syllable, and 40 for the third syllable. The number who performed worse with a contralateral interrupter than no interrupter was 38 for the first syllable, 41 for the second syllable, and 40 for the third syllable. Importantly, the effect of the interrupter does not seem to differ whether it is ipsilateral or contralateral to the target stream.

Figure 3(C) compares the performance degradation caused by ipsilateral interrupters and contralateral interrupters at the group level. Figure 3(D) plots these data for individual participants; here, solid lines indicate that the interrupter had a greater effect on the second syllable than the first syllable (left line segments) or third syllable (right line segments), while dashed lines indicate a smaller effect of the interrupter on syllable 2. Most participants show a larger effect of the interrupter for syllable 2 than for syllable 1 (38 of 45 for ipsilateral interrupter, 35 of 45 for

contralateral interrupter) and for syllable 3 (28 of 45 for ipsilateral interrupter, 28 of 45 for contralateral interrupter).

T-tests showed that, as with Experiment 1, performance for interrupted trials was significantly degraded for all syllables, both for the contralateral interrupter (1st syllable: $t_{44} = 5.930, p < 0.001$; 2nd syllable: $t_{44} = 11.016, p < 0.001$; 3rd syllable: $t_{44} = 7.185, p < 0.001$) and the ipsilateral interrupter (1st syllable: $t_{44} = 3.272, p = 0.006$; 2nd syllable: $t_{44} = 9.204, p < 0.001$; 3rd syllable: $t_{44} = 7.223, p < 0.001$). Subsequent repeated-measures two-way ANOVA on the difference data indicates the influence of the interrupter varies with syllable position ($F_{2,88} = 29.551, p < 0.001$) but not with interrupter direction ($F_{1,44} = 0.482, p = 0.491$); moreover, the interaction between syllable position and interrupter direction was not significant ($F_{2,88} = 0.052, p = 0.949$). *Post hoc* Tukey tests showed that the interrupter effect was larger on the 2nd syllable than on both the 1st ($p < 0.001$) and the 3rd syllables ($p = 0.015$), and larger on the 3rd syllable than the 1st syllable ($p < 0.001$).

C. Experiment 3

Experiment 3 investigated whether the effect of the interrupter depends on its timing relative to the target syllables. The early interrupter condition played the interrupter 300 ms before the onset of the first target syllable from 90° from the side contralateral to the target (“Early90”). We also tested two cases in which the interrupter occurred at the expected time of the third distractor syllable (replacing that syllable), which was 300 ms after the onset of the third target syllable. In one of these late interrupter conditions, we played the interrupter not only at the expected time but at the expected location of the third distractor syllable, from 30° in the hemifield contralateral to the target (“Late30”); we compared this to a case when it came at that same time, but from 90° in the hemifield contralateral to the target (“Late90”). We hypothesized that if a late arriving interrupter had an effect, it might be reduced if the interrupter occurred at both the time and place of an expected distractor syllable (Late30), as it might be more effectively suppressed than when it came from a novel location (Late90). Each of these conditions made up one third of the interrupted trials (Fig. 4).

We found that the preceding interrupter (Early90) had a small effect, but that there was no disruptive effect of either of the later interrupters (Late30 or Late90; see Fig. 4A)—even though these late interrupters were *closer* in time to the final target syllable than the interrupter in Experiments 1 and 2 was to the first syllable, which showed “backwards in time” effects of the interrupter.

Nine one-tailed one-sample *t* tests with Bonferroni correction were performed to determine if there was an effect of each of the interrupters on each syllable. Only the Early90 interrupter caused any significant degradation in target recall, and this effect only was present for the first syllable ($t_{39} = 3.812, p = 0.002$); neither of the late interrupters degraded target recall. Since only one type of interrupter

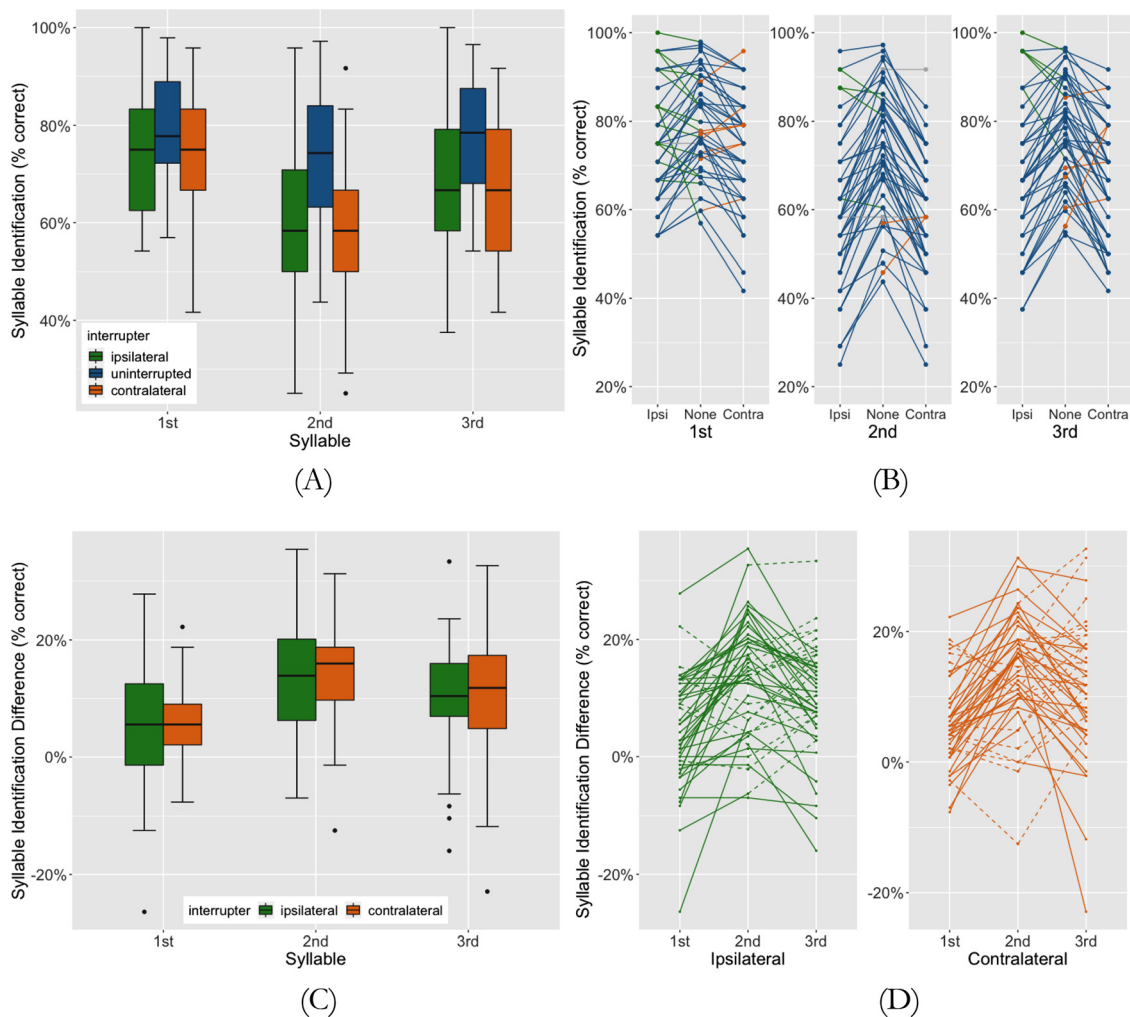


FIG. 3. (Color online) Results for Experiment 2 ($N=45$). (A) Group-level raw percent correct recall for each of the syllables when they are interrupted by an ipsilateral interrupter (green), not interrupted (blue), or interrupted with a contralateral interrupter (orange). Box plots cover from 25th to 75th percentile of performance, while range is shown by whiskers. Outliers are shown by asterisks. (B) Raw percent correct recall plotted for individual participants for each of the syllables. In each panel, line segments connect individual participants' scores when there is an ipsilateral interrupter and no interrupter (left segments) and when there is no interrupter and a contralateral interrupter (right segments). The color of each line segment denotes which condition leads to better performance for that segment: blue segments correspond to cases with better performance for the uninterrupted condition (blue), while green and orange denote better performance for the ipsilateral and contralateral interrupters than for no interrupter, respectively. (C) Group-level within-subject difference in percent recall, showing the decrement in performance caused by the interrupter for each syllable relative to performance in the uninterrupted condition. T-tests on this group level data reveal significant effects of both the contralateral and ipsilateral interrupters on each syllable. (D) Difference in percent recall with respect to uninterrupted condition plotted for individual participants with an ipsilateral interrupter (left) and a contralateral interrupter (right). Lines connect individual differences for each syllable. The format of each line segment indicates whether for the pair of connected syllables, the effect of the interrupter was greater for syllable 2 (solid) or syllables 1 or 3 (dashed). Follow-up statistical tests confirm that the effect of the interrupter varies with syllable position, and differs significantly between all pairs of syllables (largest for syllable 2, intermediate for syllable 3, and least for syllable 1).

had an effect, and only on one syllable, no subsequent analysis was done on the difference data. Furthermore, to reduce clutter in Fig. 4(B), we show only the individual data for no interrupter and Early90 interrupter (where there was a significant effect).

D. Experiment 4

Experiment 4 was designed to study whether the effect of the interrupter was due in part to its relative novelty (Fig. 5). Conditions were identical to those used in Experiment 1, however, the interrupted trials made up half of the trials in this experiment; thus, on any given trial, the

interrupter was just as likely to occur as to not occur. As in Experiment 1, the interrupter impaired performance on the second syllable, which was presented right after the interrupter [Fig. 5(A); blue for uninterrupted, orange for interrupted], however, the effect on the first and third syllable was smaller than in Experiment 1. Figure 5(B) shows, for each individual, whether performance was better in uninterrupted (blue line segments) or interrupted (orange line segments) trials. Although a majority of participants were better for uninterrupted than interrupted trials (29 of 45 for syllable 1, 32 for syllable 2, and 30 for syllable 3), the percentage of participants showing a degradation due to the interrupter was smaller than in Experiment 1. Figure 5(C) plots the effect of

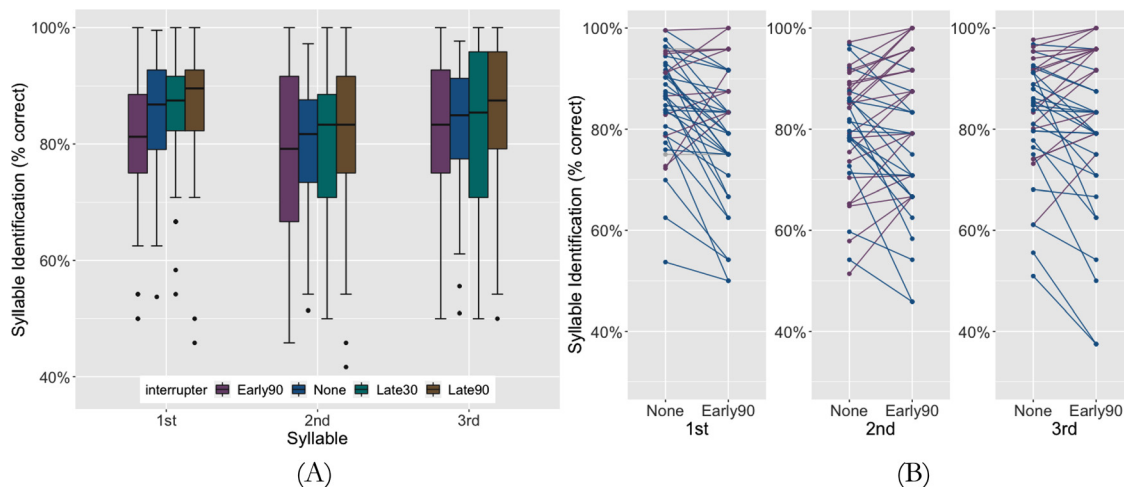


FIG. 4. (Color online) Results for Experiment 3 ($N = 40$) (A) Group-level raw percent correct recall for each of the syllables when they are uninterrupted (blue), interrupted with the Early90 interrupter (purple), Late30 interrupter (cyan), and Late90 interrupter (brown). Box plots cover from 25th to 75th percentile of performance, while range is shown by whiskers. Outliers are shown by asterisks. (B) Raw percent correct recall for individual participants for each of the syllables with no interrupter and with the Early90 interrupter. Late interrupted conditions are not shown since they caused no significant effect on target recall. The line color indicates whether, for that syllable, performance for a given subject was better for the uninterrupted condition (blue) or for the early interrupted condition (purple).

the interrupters at the group level, while Fig. 5(D) plots these differences for individual subjects for each syllable, using the same plotting scheme as in Fig. 2(D).

Three one-tailed one-sample t -tests revealed the interrupter had a significant effect on the second syllable ($t_{44} = 3.307$, $p = 0.003$); however, unlike in Experiment 1, the interrupter caused no significant effect on either the first or third syllables. Given this, no follow-up ANOVA was performed.

IV. DISCUSSION

The results of the current study demonstrate that in a spatial selective attention task, a salient interrupter presented either just before or during a target stream impairs the ability to recall target syllables. Thus, this paradigm demonstrates the competition between top-down, endogenous attention and bottom-up attention driven by salient auditory events.

A. The interrupter has the greatest effect on target syllables that came after it

In Experiments 1, 2, and 4, the interrupter started after the initial target syllable and before the second target syllable. In these experiments, we hypothesized that the interrupter would disrupt selective attention to the target, leading to poorer accuracy in recalling the subsequent syllables. In addition to disrupting attention, the interrupter also overlapped temporally with the first 125 ms of the second syllable, which may have resulted in some energetic masking that could have contributed to making it difficult to identify that syllable. Results supported this: when the interrupter came before the second target syllable, the interrupter significantly interfered with recall of the second syllable for Experiments 1, 2, and 4, as well as of the third syllable for

Experiments 1 and 2. Moreover, ANOVA on the size of the interference confirmed that in both Experiments 1 and 2, the interference for syllable two was greater than for the third syllable and the first syllable, while the interference on syllable three was greater than the interference on the first syllable. Similarly, in Experiment 3 we hypothesized that the Early90 would disrupt attention for the first target syllable, which began just afterward. In this experiment, the early interrupter had a significant effect on performance for the first target syllable, but not for later syllables.

Focusing spatial selective auditory attention on one stream enhances neural responses to the stream at the attended location and suppresses responses to other, competing objects (Choi *et al.*, 2013; Hillyard *et al.*, 1998). However, attention can be involuntarily hijacked by salient stimuli (Buschman and Miller, 2007; Shinn-Cunningham, 2017). In Experiments 1 and 2, the MEOH happened only on 25% of the trials and was therefore relatively unexpected, which can add to its salience. However, even in Experiment 4, where the interrupter appears on half the trials and is thus expected, it still disrupts spatial selective attention to the target for the second target syllable. We believe that the interrupting cat MEOH grabs attention involuntarily even in Experiment 4 because it sounds completely unlike the target and distractor speech streams; thus, it is always heard as a new object, even though it was expected.

Once attention is diverted away from the target by the interrupter, it takes some amount of time to shift attention back to the target stream (Mondor and Zatorre, 1995). Estimates from previous visual and auditory attention experiments suggest that reorienting attention requires on the order of a few hundreds of ms (Larson and Lee, 2013; Logan, 2005; Shapiro *et al.*, 1997). By placing our interrupter only 125 ms before the onset of the second target syllable, we therefore expected to see a performance cost for

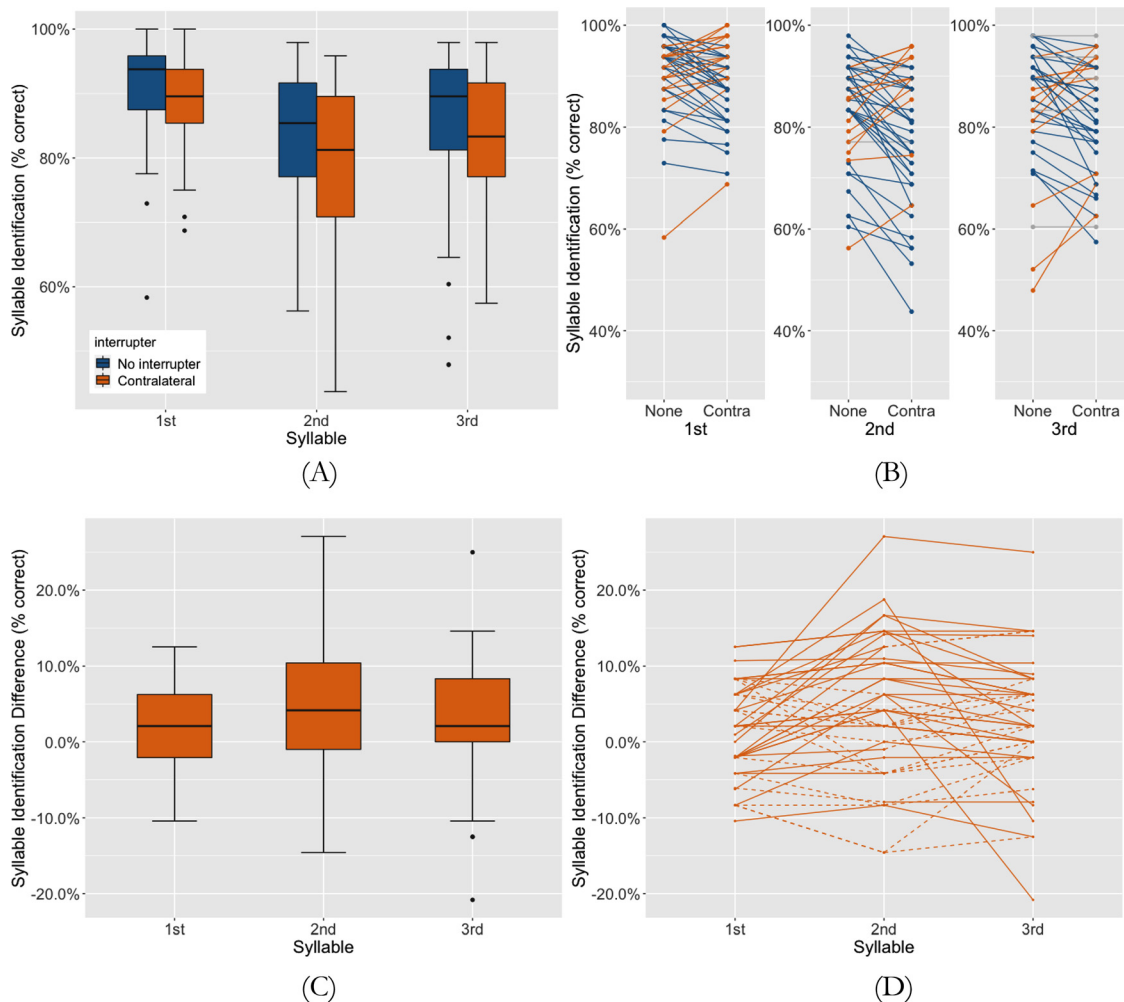


FIG. 5. (Color online) Results for Experiment 4 ($N = 45$). (A) Group-level raw percent correct recall for each of the syllables when they are uninterrupted (blue) and interrupted (orange). Box plots cover from 25th to 75th percentile of performance, while range is shown by whiskers. Outliers are shown by asterisks. (B) Raw percent correct recall plotted for individual participants for each syllable with and without the interrupter. The color of the line indicates whether the performance for a given subject is better for the uninterrupted condition (blue) or the interrupted condition (orange) for that syllable. (C) Group-level within-subject difference in percent recall, the effect of the interrupter for each syllable relative to performance in the uninterrupted condition. T-tests on the performance difference data reveal a significant negative effect of the interrupter on the second syllable. (D) Difference in percent recall with respect to uninterrupted condition plotted for individual participants. Lines connect individual differences for each syllable. The line type indicates whether for the pair of connected syllables, the effect of the interrupter was greater for syllable 2 (solid) or syllables 1 or 3 (dashed).

the ability to recall the second target syllable in Experiments 1, 2, and 4. In Experiment 3, the early interrupter began 300 ms before the first target syllable, yet still led to some degradation in recall of the initial target syllable. These results are consistent with our hypothesis that salient interrupters involuntarily grab attention, disrupting top-down spatial auditory selective attention, which then interferes with recall of the subsequent target syllable (the second target syllable in Experiments 1, 2, and 4 and the first target syllable for early interrupters in Experiment 3).

In Experiments 1 and 2, in addition to the expected drop in accuracy for the second target syllable that happened right after the MEOW, recall was also disrupted for the third target syllable, which began 725 ms after the interrupter. Based on previous reports (e.g., Larson and Lee, 2013), a delay of this duration should have been sufficient to allow listeners to reorient top-down attention before the third syllable began. In

Experiment 3, the early interrupter had no discernable impact on recall of the second syllable, which began 900 ms after the MEOW. Together, these results might be interpreted as showing that reorienting top-down attention to the target stream in our paradigm requires on the order of 700–900 ms—but this is much longer than estimates of the time it takes to reorient top-down attention in past studies.

It is worth noting this time estimate (700–900 ms) assumes that a listener would try to reorient immediately after the onset of the interrupter; however, it is also possible that they would not reorient this rapidly. The interrupter duration was 250 ms; if a listener remained focused on the interrupter until it finished playing, this estimate would be closer to 450–650 ms, which is nearer to past estimates of the time to reorient attention.

Still, we do not believe that the pattern of errors that listeners made can be explained simply by a need to reorient

top-down attention. For the majority of the participants (34 out of 45 in Experiment 1; 31 out of 45 for an ipsilateral interrupter and 38 out of 45 for a contralateral interrupter in Experiment 2; 29 out of 45 in Experiment 4), the interrupter also degraded recall accuracy for the first target syllable, which had completed playing *before* the cat sound began. Thus, as discussed later, disruption of top-down attention and a subsequent need to reorient cannot fully account for the performance costs introduced by a salient interrupter.

B. The effect of a salient interrupter does not vary with its location

As laid out in Sec. I, the time it takes to reorient top-down attention grows as the spatial separation between auditory objects increases (Mondor and Zatorre, 1995; Rhodes, 1987), which suggests that an interrupter from the opposite hemifield might cause a longer-duration and larger disruption in attention. Alternatively, top-down spatial attention suppresses distracting streams more effectively when the spatial separation from the target stream is greater (Best *et al.*, 2006; Best *et al.*, 2008). This might suggest that top-down spatial attention might be more effective at suppressing an interrupter that comes from the hemifield opposite the attended target stream. The current results do not support either of these ideas. Instead, Experiment 2 showed that the interrupter had essentially identical effects whether it appeared contralateral to or ipsilateral to the target stream: there were no significant differences in the effect of interrupters from opposite hemifields. This finding shows that rather than depending on the spatial separation between the target stream and interrupter, the disruption caused by a salient new sound operates independently of top-down spatial attention. Of course, it is possible that changes in the spatial separation between the interrupter and the attended target produced opposing effects of roughly the same size, leading to no net effect. Still, it is more parsimonious to assume that spatial configuration did not influence the disruption caused by the interrupter in our experiments.

It is worth noting that in our paradigm, target and distractor streams were male speech, which shares little semblance to the interrupter, a cat MEOW. Given the dissimilarity of the interrupter and the other streams within each trial, the interrupter undoubtedly always was heard as a distinct, new stream. Further, in Experiments 1 and 2, the interrupter was infrequent and unexpected, occurring only in 25% of the trials. Also, we ensured that the interrupter was salient by playing it at a level 7–10 dB higher than the target and distractor syllables. All of these factors contribute to making the interrupter highly salient. It may be that in this kind of situation, where the interrupter is highly salient no matter what its spatial location, the interrupter location has no influence on performance. It could be that an interrupter that is less salient would have a more nuanced effect on performance, which might reveal a spatial dependence on its effectiveness. This idea could be pursued in future studies.

It is also possible that the spatial location of the interrupter would influence how much it disrupts target recall if

the interrupter is similar to the target stream, such as when a male talker interrupts an attended stream of male speech. In such cases, top-down attention may suppress the interrupter, and the effectiveness of the suppression may depend on the spatial separation between the target location (where attention is focused) and the interrupter. In fact, even semantic similarity of the target and interrupter might influence the effectiveness of the interrupter; spatial separation has been shown to interact with semantic features when listeners perform a divided attention task (McCloy and Lee, 2015). Specifically, the more similar the target and interrupter (in location, in timbre, and even in semantics), the more likely they are to be confused and perceptually entangled, rather than being perceived as separate objects. In such situations, the spatial relationship between similar streams is likely to influence how strongly an interrupter disrupts top-down spatial attention. Further experiments should explore this possibility.

C. Interruptions during a target stream interfere with storage of attended syllables

Surprisingly, in both Experiments 1 and 2 there was a significant effect of the interrupter on performance for the first syllable, which finished playing before the interrupter began. Presumably, listeners had been selectively attending to the target, as instructed, prior to the occurrence of the interrupter. Thus, this backwards-in-time influence suggests that the interrupter interfered not only with focusing attention on target syllables that occurred after the interrupter but also with storing the already-attended first target syllable in working memory.

Working memory and attention are closely related (Gazzaley and Nobre, 2012). Specifically, attention seems to work as a “gatekeeper” for working memory; only information that is “let through” by attention can be subsequently stored in working memory (Awh *et al.*, 2006). However, several studies have shown that an attentional bottleneck alone cannot explain what information gets encoded and maintained in working memory (Lewis-Peacock *et al.*, 2018; Oberauer, 2018, 2019). Some information that is attended may be dropped, and not stored in working memory if it is not behaviorally relevant (Lewis-Peacock *et al.*, 2018; Oberauer, 2018). At other times, attention may fail to perfectly filter out irrelevant information, leading irrelevant information to interfere with relevant information that a participant wishes to save in working memory (Hakim *et al.*, 2020).

In our study, focusing attention on the first target syllable cannot have been the problem, as it finished playing before the interrupter occurred. Participants also knew that they should store the first target syllable, as it was relevant to the task; they should not have intentionally decided to not store it in working memory. Even so, the salient bottom-up interrupter interfered with recall of the initial target syllable. We hypothesize that an interrupter occurring in the middle of the three target syllables, which form a single perceptual stream, disrupted storage of that entire stream. As discussed

in Sec. IVD, this idea gains indirect support from Experiment 3, which shows no backwards-in-time effect of late interrupters that occur after the end of the entire target stream.

One thing to note here is that within each trial, the target syllables were randomly selected without replacement. Thus, on each trial, each syllable appeared exactly once in the target stream. It is quite possible that participants utilized this fact and restricted their responses to “legal” answers in which they named each syllable only once. If so, then responses to the individual syllables are not strictly independent of one another.

To explore this possibility, we undertook a *post hoc* analysis of the errors. Figure 6 shows, for each experiment and condition, a breakdown of how many responses were “repeat” errors (reporting the same syllable more than once within a trial; bottom portions of the bars), “permute errors” (reporting each of the syllables only once on a trial, but in the wrong order; middle sections of the bars), or correct answers (top portions of the bars). We found that more than half of the participants in each experiment made repeat errors at least once (68.9% in Experiment 1, 62.2% in Experiment 2, 52.5% in Experiment 3, 55.6% in Experiment 4). However, overall, the number of repeat errors was small and did not change across conditions. In contrast, the number of permute errors was larger and generally increased in conditions with the interrupter compared to those with no interrupter.

The high likelihood of permutation errors could arise because listeners were sure they heard each syllable, but were confused about the order of presentation, or because they were directly influenced by the expectation that each syllable only was presented once per trial. We have no way of separating these two possibilities. Moreover, the limited number of permute errors made in these experiments makes it impossible to do a meaningful finer-grain analysis of the error patterns. Future experiments in which the target stream is not constrained to include each syllable only once could be conducted to illuminate whether listeners’ tendency to make permute errors in the current experiment reflects confusion about syllable order, rather than a cognitive strategy of restricting answers to permutations of target syllable order.

D. Storage of already-attended syllables was only affected when an interruption occurred during the target stream

Based on results from Experiments 1 and 2, where recall of the first target syllable was disrupted by the subsequent interrupter, we thought that late interrupters *might* interfere with recall of the final target syllable in Experiment 3. Instead, we did not see any effect of the late interrupters on recall of any of the target syllables in Experiment 3. It is worth noting that the timing between the onsets of the final target syllable and the late interrupter in Experiment 3 (300 ms), where there was no effect, was even shorter than the delay between the first target syllable and the interrupter in Experiments 1 and 2 (475 ms), where the recall of the first syllable was disrupted. Given this, one might have expected an even greater effect of the late interrupters in Experiment 3 on the final target syllable than the effect of the interrupter on recall of the first syllable in Experiments 1 and 2. This discrepancy suggests that the interferer only has a strong backwards-in-time effect on recall when it occurs in the middle of the target stream. This kind of disruption hints that the target stream is normally processed and stored in memory as a single object, not as separate syllables. This idea could be tested by creating a target stream that is not perceived as a single stream but is instead heard as distinct events—for instance, by changing the talker from target syllable to target syllable (Carter *et al.*, 2019; Lim *et al.*, 2021). If the individual target utterances are heard and stored as separate items, an interrupter should not have the same impact on items that were heard and stored before the interrupter occurred.

E. An interruption during a target stream has larger effects than does an interrupter that precedes the target stream

To better compare results across the different experiments, we computed the Cohen’s *d* effect size of the interrupters on performance for each of the syllables. This analysis, shown in Table I, reinforces the idea that a disruption occurring in the middle of an ongoing stream is, in fact, qualitatively more disruptive than a salient event that does

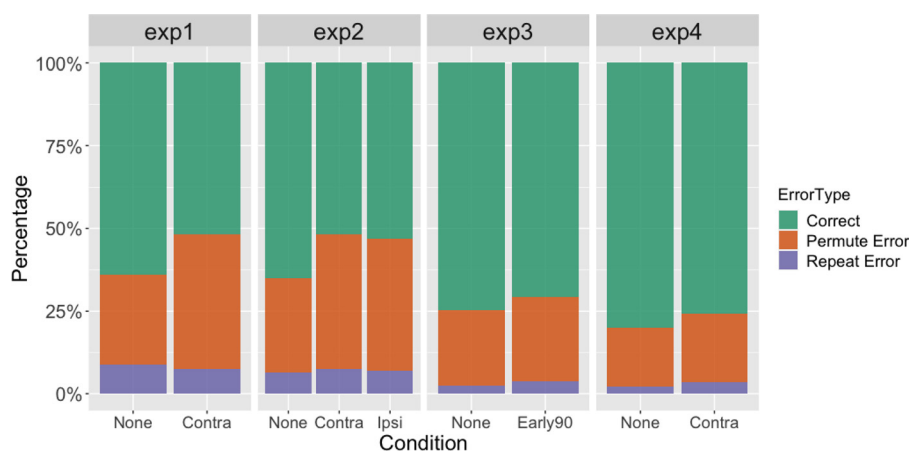


FIG. 6. (Color online) Error analysis for the no interrupter condition and interrupted conditions in which the interrupter had a statistically significant effect for each experiment. Each bar breaks down all trials to show the mean percentage of correct trials (green), trials with permutation errors (orange), and trials with repeat errors (lavender).

TABLE I. Effect sizes (Cohen's d) of the interrupters on each syllable for Experiments 1, 2, and 4.

Experiment	Syllable 1	Syllable 2	Syllable 3
Exp1 (contralateral)	0.54 (medium)	1.36 (large)	1.04 (large)
Exp2 (contralateral)	0.51 (medium)	1.00 (large)	0.86 (large)
Exp2 (ipsilateral)	0.39 (small)	0.85 (large)	0.70 (medium)
Exp3 (Early90)	0.51 (medium)	0.07 (negligible)	0.17 (negligible)
Exp3 (Late90)	0.12 (negligible)	0.06 (negligible)	0.20 (small)
Exp3 (Late30)	0.03 (negligible)	0.04 (negligible)	0.07 (negligible)
Exp4 (contralateral)	0.19 (negligible)	0.38 (small)	0.25 (small)

not occur in the middle of the target. Specifically, in Experiments 1 and 2, the interrupter not only causes a large-size effect on recall of syllable 2 but also a medium- to large-size (depending on the exact experiment and condition) effect on syllable 3, which begins a full 725 ms after the interrupter. In contrast, in Experiment 3 the early interrupter, which happens only 300 ms before the first target syllable, shows a smaller effect size on syllable 1 (0.51) than the effects on syllables 2 and 3 in any of the conditions in Experiments 1 and 2 (0.70–1.36). That is, an interrupter that occurs during the presentation of the target stream (Experiments 1 and 2) both affects recall of a syllable that finished playing before the interrupter occurred (the first syllable) and has a larger effect on recall of subsequent syllables, at later delays, than the impact of an interrupter that occurs before the start of a target stream (early interrupter in Experiment 3). Together, these results argue that an interrupter during an ongoing stream interferes not only with attentional focus on the target stream but with storage of that stream in memory.

The hypothesis that a salient interruption during a target stream is especially disruptive to storing the target in working memory can be tested in future experiments by increasing the number of items in the target stream and seeing how many preceding target items are disrupted by a later interferer. Moreover, electroencephalography (EEG) studies could be conducted to track effects of the interrupter on responses to the target syllables and on working memory load change before and after the interruption.

F. Interruption effect decrease with increased likelihood of interruption

Experiment 4 was almost identical to Experiment 1, except it has an increased likelihood of interruptions. Specifically, each of the trials was equally likely to be interrupted or not. This more frequent and more “expected” interrupter had a smaller effect than that caused by otherwise identical contralateral interrupters in Experiments 1 or 2. In fact, the interrupter in Experiment 4 did not even have a significant effect on recall of the first or the third syllables. Table I shows that for all three syllables, the interrupter had a smaller sized effect in Experiment 4 than the comparable contralateral interrupters in the first two experiments.

The decrease in interruption effect size aligns with our expectations and is consistent with previous studies showing

that behavioral distraction arises from a violation of expectation based on learned conditional probabilities of events (Nösl *et al.*, 2012; Parmentier *et al.*, 2011; Vachon *et al.*, 2012). Our participants were given no explicit instructions to anticipate an interrupting MEOW sound (which might attenuate the interruption effect; see Röer *et al.*, 2015). However, they were able to build up an expectation of whether or not a trial was likely to be interrupted from the frequency of interruptions over previous trials. In Experiment 4, with the MEOW occurring on half of all trials, participants appear to expect an interruption, making it less surprising and less salient, and reducing its impact on target recall.

One caveat is important to note, however: overall performance in uninterrupted conditions was better in Experiment 4 (mean accuracy: 86.25%) than Experiments 1 or 2 (mean accuracy: 75.35% and 76.77%, respectively). If baseline performance is sufficiently high, it may limit the measured impact of the interruption on percent correct recall of the target stream and thus explain the smaller effect of the interrupter in Experiment 4. Such a change in baseline performance could be due to a few different factors. First, we used different syllable recordings in the final experiment; specifically, we re-recorded the stimuli used in Experiment 4 to reduce syllable confusion. Additionally, differences in the subject groups might contribute to differences in baseline performance across tasks. Although we recruited participants for all of the experiments using identical procedures and used relatively large subject group sizes, intersubject differences in performance are fairly pronounced [e.g., see Figs. 2(B), 3(B), 4(B), and 5(B)]. Given previous evidence that expected events are less salient than unexpected events, our intuition is that the main reason that the interrupters in Experiment 4 lead to smaller impacts on target recall is due to the interrupter frequency, not ceiling effects on overall performance; however, further experiments are needed to confirm this.

V. CONCLUSION

Salient bottom-up interrupters degrade recall of target streams during a top-down spatial selective auditory attention task. The location of the interrupter has no statistically significant effect on its impact, suggesting a mechanism that operates independent of top-down spatial filtering. An unexpected interrupter degrades recall of the entire stream that is being interrupted, not only the subsequent target syllable, while a more expected interrupter has smaller effects that are only statistically significant on the immediately subsequent syllable. These results suggest that an unexpected interrupter interferes not only with focusing selective attention but also with storage of the attended target stream in working memory.

ACKNOWLEDGMENTS

This work was supported by the Montgomery Research Fellow Fund from the CMU Neuroscience Institute and grants from the National Institute on Deafness and Other Communication Disorders (R01DC019126 to BGSC and R21DC018408 to CB).

- Allen, K., Alais, D., and Carlile, S. (2009). "Speech intelligibility reduces over distance from an attended location: Evidence for an auditory spatial gradient of attention," *Percept. Psychophys.* **71**, 164–173.
- Arbogast, T. L., and Kidd, G., Jr. (2000). "Evidence for spatial tuning in informational masking using the probe-signal method," *J. Acoust. Soc. Am.* **108**, 1803–1810.
- Awh, E., Vogel, E. K., and Oh, S.-H. (2006). "Interactions between attention and working memory," *Neuroscience* **139**, 201–208.
- Best, V., Gallun, F. J., Ihlefeld, A., and Shinn-Cunningham, B. G. (2006). "The influence of spatial separation on divided listening," *J. Acoust. Soc. Am.* **120**, 1506–1516.
- Best, V., Ozmeral, E. J., Kopco, N., and Shinn-Cunningham, B. G. (2008). "Object continuity enhances selective auditory attention," *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13174–13178.
- Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), p. 800.
- Broadbent, D. E. (1954). "The role of auditory localization in attention and memory span," *J. Exp. Psychol.* **47**, 191–196.
- Brown, C. A. (2014). "Binaural enhancement for bilateral cochlear implant users," *Ear Hear.* **35**, 580–584.
- Buschman, T., and Miller, E. (2007). "Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices," *Science* **315**, 1860–1862.
- Carter, Y. D., Lim, S.-J., and Perrachione, T. K. (2019). "Talker continuity facilitates speech processing independent of listeners' expectations," in *Proceedings of the 19th International Congress of Phonetic Sciences*, August 5–9, Melbourne, Australia.
- Chen, Z. (2012). "Object-based attention: A tutorial review," *Atten. Percept. Psychophys.* **74**, 784–802.
- Cherry, E. C. (1953). "Some experiments on the recognition of speech, with one and with two ears," *J. Acoust. Soc. Am.* **25**, 975–979.
- Choi, I., Rajaram, S., Varghese, L., and Shinn-Cunningham, B. (2013). "Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography," *Front. Hum. Neurosci.* **7**, 115.
- Deng, Y., Choi, I., Shinn-Cunningham, B., and Baumgartner, R. (2019). "Impoverished auditory cues limit engagement of brain networks controlling spatial selective attention," *Neuroimage* **202**, 116151.
- Desimone, R., and Duncan, J. (1995). "Neural mechanisms of selective visual attention," *Ann. Rev. Neurosci.* **18**, 193–222.
- Duncan, J. (1984). "Selective attention and the organization of visual information," *J. Exp. Psychol. Gen.* **113**, 501–517.
- Duncan, J. (2006). "EPS mid-career award 2004: Brain mechanisms of attention," *Q. J. Exp. Psychol.* **59**, 2–27.
- Folk, C. L., and Remington, R. W. (2015). "Unexpected abrupt onsets can override a top-down set for color," *J. Exp. Psychol. Hum. Percept. Perform.* **41**, 1153–1165.
- Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). "Auditory attention—Focusing the searchlight on sound," *Curr. Opin. Neurobiol.* **17**, 437–455.
- Gazzaley, A., and Nobre, A. C. (2012). "Top-down modulation: Bridging selective attention and working memory," *Trends Cogn. Sci.* **16**, 129–135.
- Goldstein, E. B., and Fink, S. I. (1981). "Selective attention in vision: Recognition memory for superimposed line drawings," *J. Exp. Psychol. Hum. Percept. Perform.* **7**, 954–967.
- Hakim, N., Feldmann-Wüstefeld, T., Awh, E., and Vogel, E. K. (2020). "Perturbing neural representations of working memory with task-irrelevant interruption," *J. Cogn. Neurosci.* **32**, 558–569.
- Hillyard, S. A., Vogel, E. K., and Luck, S. J. (1998). "Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence," *Philos. Trans. R Soc. London, B* **353**, 1257–1270.
- Huang, N., and Elhilali, M. (2020). "Push-pull competition between bottom-up and top-down auditory attention to natural soundscapes," *Elife* **9**, e52984.
- Kaya, E. M., and Elhilali, M. (2014). "Investigating bottom-up auditory attention," *Front. Hum. Neurosci.* **8**, 327.
- Kidd, G., Jr., Mason, C. R., and Gallun, F. J. (2005). "Combining energetic and informational masking for speech identification," *J. Acoust. Soc. Am.* **118**, 982–992.
- Larson, E., and Lee, A. K. C. (2013). "Influence of preparation time and pitch separation in switching of auditory attention between streams," *J. Acoust. Soc. Am.* **134**, EL165–EL171.
- Lewis-Peacock, J. A., Kessler, Y., and Oberauer, K. (2018). "The removal of information from working memory," *Ann. N.Y. Acad. Sci.* **1424**, 33–44.
- Lim, S.-J., Carter, Y. D., Michelle Njoroge, J., Shinn-Cunningham, B. G., and Perrachione, T. K. (2021). "Talker discontinuity disrupts attention to speech: Evidence from EEG and pupillometry," *Brain Lang.* **221**, 104996.
- Logan, G. D. (2005). "The time it takes to switch attention," *Psychon. Bull. Rev.* **12**, 647–653.
- Marinato, G., and Baldauf, D. (2019). "Object-based attention in complex, naturalistic auditory streams," *Sci. Rep.* **9**, 2854.
- McCloy, D. R., and Lee, A. K. C. (2015). "Auditory attention strategy depends on target linguistic properties and spatial configuration," *J. Acoust. Soc. Am.* **138**, 97–114.
- Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., and Chait, M. (2021). "An online headphone screening test based on dichotic pitch," *Behav. Res.* **53**, 1551–1562.
- Mondor, T. A., and Zatorre, R. J. (1995). "Shifting and focusing auditory spatial attention," *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 387–409.
- Müller, H. J., Geyer, T., Zehetleitner, M., and Krummenacher, J. (2009). "Attentional capture by salient color singleton distractors is modulated by top-down dimensional set," *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 1–16.
- Murdock, B. B., Jr. (1962). "The serial position effect of free recall," *J. Exp. Psychol.* **64**, 482–488.
- Neisser, U., and Becklen, R. (1975). "Selective looking: Attending to visually specified events," *Cogn. Psychol.* **7**, 480–494.
- Nösl, A., Marsh, J. E., and Sörqvist, P. (2012). "Expectations modulate the magnitude of attentional capture by auditory events," *PLoS One* **7**, e48569.
- Oberauer, K. (2018). "Removal of irrelevant information from working memory: Sometimes fast, sometimes slow, and sometimes not at all," *Ann. N.Y. Acad. Sci.* **1424**, 239–255.
- Oberauer, K. (2019). "Working memory and attention—A conceptual analysis and review," *J. Cogn.* **2**, 36.
- Parmentier, F. B. R., Elsley, J. V., Andrés, P., and Barceló, F. (2011). "Why are auditory novels distracting? Contrasting the roles of novelty, violation of expectation and stimulus change," *Cognition* **119**, 374–380.
- Rhodes, G. (1987). "Auditory attention and the representation of spatial information," *Percept. Psychophys.* **42**, 1–14.
- Rock, I., and Gutman, D. (1981). "The effect of inattention on form perception," *J. Exp. Psychol. Hum. Percept. Perform.* **7**, 275–285.
- Röer, J. P., Bell, R., and Buchner, A. (2015). "Specific foreknowledge reduces auditory distraction by irrelevant speech," *J. Exp. Psychol. Hum. Percept. Perform.* **41**, 692–702.
- Salmi, J., Rinne, T., Koistinen, S., Salonen, O., and Alho, K. (2009). "Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention," *Brain Res.* **1286**, 155–164.
- Schroeder, M. (1965). "New method of measuring reverberation time," *J. Acoust. Soc. Am.* **37**, 409–412.
- Shapiro, K. L., Raymond, J. E., and Arnell, K. M. (1997). "The attentional blink," *Trends Cogn. Sci.* **1**, 291–296.
- Shinn-Cunningham, B. (2008). "Object-based auditory and visual attention," *Trends Cogn. Sci.* **12**, 182–186.
- Shinn-Cunningham, B. (2017). "Cortical and sensory causes of individual differences in selective attention ability among listeners with normal hearing thresholds," *J. Speech. Lang. Hear. Res.* **60**, 2976–2988.
- Shinn-Cunningham, B., Best, V., and Lee, A. K. C. (2017). "Auditory object formation and selection," in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay (Springer International Publishing, Cham), pp. 7–40.
- Sussman, E. S. (2017). "Auditory scene analysis: An attention perspective," *J. Speech. Lang. Hear. Res.* **60**, 2989–3000.
- Vachon, F., Hughes, R. W., and Jones, D. M. (2012). "Broken expectations: Violation of expectancies, not novelty, captures auditory attention," *J. Exp. Psychol. Learn. Mem. Cogn.* **38**, 164–177.