Object-based suppression in auditory selective attention: The influence of statistical learning

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of The Ohio State University

By

Heather Daly, M. A.

Graduate Program in Psychology

The Ohio State University

2019

Thesis Committee

Mark Pitt, Advisor

Andrew Leber

Alexander Petrov

Abstract

Many definitions of selective attention tend to reference two components: a facilitatory mechanism that enhances the signal of interest, and an inhibitory mechanism that suppresses irrelevant and potentially distracting signals. These mechanisms have been studied extensively in vision, but less is known about their operation in audition. The present investigation tested whether suppression in auditory selective attention is sensitive to statistical regularities in complex scenes, which was shown in recent work in vision (Wang & Theeuwes, 2018). Participants listened to complex scenes consisting of several voices saying series of numbers and a distracting environmental noise. There were two possible distracting noises, one of which occurred much more frequently (70%) than the other (30%). One voice on each trial was a gender singleton, and participants were instructed to find that voice and report whether it was saying even or odd numbers. If suppression is an active component of auditory selective attention, it should reduce the influence of the distracting noise that occurs more frequently. Results revealed significantly faster RTs when the high-probability distracting noise was in the scene relative to when the low-probability distracting noise was in the scene, suggesting that participants used the frequency of the distractor across trials to aid performance. This result demonstrates that suppression mitigates the detrimental influence of a frequently occurring distracting sound.

Vita

2014.....B.S. Psychology, Ball State University

2017.....M.A. Experimental Psychology, James Madison University

2018 to present.....Graduate Teaching Associate, Department of Psychology, The Ohio State University

Publications

Daly, H.R., & Hall, M.D. (2018). Not all musicians are created equal: Statistical concerns regarding the categorization of participants. *Psychomusicology: Music, Mind, and Brain, 28*(2), 117-126.

Fields of Study

Major Field: Psychology

Table of Contents

List of Figures

Introduction

Everyday life requires listeners to focus on a sound of interest while ignoring simultaneous competing sounds. This can include listening to a specific voice during a conversation in a noisy room, trying to hear your own name in a crowed waiting room, or even focusing on a specific instrument when listening to a piece of music. Auditory selective attention underlies all of these abilities, and has been a topic of interest since the 1950s when the "cocktail party problem" was first introduced (Cherry, 1953). The cocktail party problem refers to our ability to listen to a specific voice in the presence of simultaneous auditory input. This is a requirement of everyday life, but the processes underlying this ability have remained elusive.

Selective attention is thought to consist of two mechanisms: a facilitatory mechanism that enhances the signal of interest, and an inhibitory mechanism that suppresses potentially distracting signals (e.g., see Alain & Bernstein, 2008; Bressler, Masud, Bharadwaj, & Shinn-Cunningham, 2014; Gazzaley, Cooney, McEvoy, Knight, & D'Esposito, 2005). These mechanisms have been well-established in vision, but relatively less is known about their operation, characteristics, and interactions in audition. Of the two mechanisms, enhancement has been studied in greater depth, with research suggesting that selective attention is improved by target voice continuity (Bressler et al., 2014; Samson & Johnsrude, 2016), target location continuity (Best, Ozmeral, Kopčo, & Shinn-Cunningham, 2008), and by statistical regularities in target location (Addleman & Jiang, 2019). Despite these advances in understanding enhancement, there has been less focus on understanding the characteristics of suppression in auditory selective attention.

Klein and Stolz (2015) provided compelling evidence in favor of a suppression mechanism in auditory attention by implementing a rapid serial auditory presentation (RSAP) task to look at intertrial repetition effects. Participants heard 12 sequentially-presented vowel sounds on each trial and were instructed to listen for the vowel that was different from the other 11 vowels (oddball task). Each vowel was randomly presented to either the left or right ear, and participants were instructed to report the ear in which the oddball vowel appeared. Error rates and response times were the highest when a previously-irrelevant vowel sound became the target oddball on a subsequent trial, indicating that participants had suppressed that sound on the previous trial.

Nolden, Ibrahim, and Koch (2019) provided additional evidence of suppression by demonstrating how increased preparation time before the onset of an auditoryscene differentially influenced target and distractor processing. Each scene consisted of two dichotically-presented numbers: one spoken by a female voice, and one spoken by a male voice. An auditory cue tone presented 400ms or 1200ms before the onset of the scene let participants know which gender was to be attended. Participants were instructed to judge whether the number spoken by the target gender was greater than five or less than five. To look at specific effects associated with target enhancement and distractor suppression, the distracting voice appeared 200ms before or after the target voice. The logic of this manipulation was based on the assumption that the stimulus presented first is also the first to be processed, so effects in the target-first and distractor-first conditions should reflect target and distractor processing, respectively. If preparation time was able to improve target enhancement, then RTs were expected to be faster when the target was presented first after a long cue-target interval (CTI) relative to a short CTI.

If preparation time was beneficial for distractor suppression, then a similar pattern was expected when the distractor was presented first. The researchers observed a benefit of increased preparation time in both situations, but found the largest RT decrease when the distractor came before the target, suggesting that preparation time benefitted distractor suppression more than it benefitted target enhancement. The selective benefit observed for distractor suppression indicates that suppression is a fundamental process in auditory selective attention.

Melara, Tong, and Rao (2012) explored whether auditory suppression abilities could be trained. They implemented a distractor suppression training program to investigate how improved suppression abilities influenced selective attention. Participants practiced detecting a target sound among a stream of standard sounds in one ear while simultaneously ignoring a distracting sound in the other ear that gradually increased in intensity. The training led to faster and more accurate target detection in a post-test, which the researchers attributed to participants' improvement in distractor suppression abilities as a result of training. This result suggests that an individual's ability to suppress distracting sounds is not fixed, and can likely improve with practice.

Suppression in auditory selective attention has not only been demonstrated for objects and features, but also for locations. Allen, Alais, Shinn-Cunningham, and Carlile (2011) showed reduced target phoneme identification performance in masker noise when the speech masker was presented from an unexpected location relative to when it was presented from an expected location. Additionally, target phoneme identification was worse when the target came from the expected masker location, providing further evidence that the location that frequently contained the masking syllable was being suppressed. In addition to demonstrating that suppression can

operate on locations in short, two-voice scenes, these results also suggest that suppression is sensitive to statistical regularities in distractor location.

Together, these studies indicate that suppression is a flexible attention mechanism that is capable of operating on a number of physical dimensions (e.g. objects, features, locations), and is likely sensitive to statistical regularities in the environment. The goal of the current investigation was to delve further into the characteristics of the suppression mechanism in auditory selective attention and determine whether it is sensitive to statistical regularities in more complex auditory scenes. For suppression to be maximally useful, it should be able to detect and utilize statistical regularities in the environment. Such a mechanism could reduce listening effort in challenging environments by learning which distracting sounds convey no useful information, and then subsequently attenuating the representations of those sounds to help the listener more efficiently focus on task-relevant information. For instance, it can take a great deal of effort to focus on what a friend is saying in a noisy coffee shop because the speech signal must compete with a number of other sounds such as coffee grinders, blenders, background music, and other voices. If a suppression mechanism could learn to ignore these extraneous sounds, it would then become much easier for the listener to remain focused on their friend's speech.

Previous research suggests that suppression of an auditory location benefits from statistical learning (Allen et al., 2011). It is therefore reasonable to suppose that suppression of an auditory object may also adapt to statistical regularities in the environment. The suppression mechanism could learn which objects are present in multi-sound scenes and use that information to attenuate the representations of distracting objects. Although it is useful to suppress locations, listeners more often focus their attention on particular objects rather than specific locations, so it

is likely more beneficial to suppress the objects themselves. At times multiple objects can come from the same perceived location, but the listener only wants to focus on one of those objects. Returning to the coffee shop example, the sound of your friend's voice could come from the same direction as the barista's voice and the loudspeaker that is playing the music. In such a situation, being able to suppress those other objects could facilitate focusing on your friend's voice despite the lack of distinct locations for each object.

Long, four-sound scenes were used in the present investigation to better reflect the temporal nature of auditory attention and to determine whether suppression could be observed in scenes that more closely captured the challenges of sustained listening. Previous work has studied auditory attention using sequential sound presentation (e.g. Klein & Stolz, 2015), or simultaneous presentation of brief sounds (e.g. Allen, et al., 2011, Melara et al., 2012; Nolden et al., 2019). This earlier work has provided promising evidence for suppression, but such simple scenarios limit what can be learned about suppression because they rarely occur in natural listening environments. Listening often involves extended attention in multi-sound environments over a period of time, but previous work has either presented a single object for a longer amount of time or several objects for a brief amount of time. It is therefore unclear whether the mechanisms underlying auditory selective attention continue to operate in similar ways in more complex environments. Because auditory scenes evolve over time, it is likely that suppression in auditory selective attention has a similar temporal nature. To better understand the role of suppression and its sensitivity to statistical regularities in auditory environments, it is important to use stimuli that allow suppression to evolve over time.

To address these goals, the current study adapted a paradigm from a visual attention study by Wang & Theeuwes (2018) that implemented statistical regularities across the course of the experiment. Their results showed that locations that contained the salient visual distractor more frequently were suppressed relative to locations that rarely contained the distractor. Although previous work has demonstrated that auditory attention can operate on locations (e.g. see Allen et al., 2011; Ihlefeld & Shinn-Cunningham, 2008; Koch & Lawo, 2014; Lewald, Hanenberg, & Getzmann, 2016), other evidence suggests that auditory attention might more naturally operate on objects (Alain & Arnott, 2000; Shinn-Cunningham, 2008; Zimmerman et al., 2016). Paralleling Wang & Theeuwes (2018), the present investigation looked at whether a distracting sound that occurred more frequently would be suppressed relative to a distracting sound that occurred less frequently, irrespective of location.

Participants were asked to listen to seven-second scenes consisting of four sounds each (four voices or three voices plus a distracting sound), find the voice that was a gender singleton, and report whether that voice was saying even or odd numbers. This attention-demanding task required listeners to search the scene for the target voice and analyze what it was saying, all while ignoring a distracting environmental sound. The task not only simulated the complexity of the listening situations we experience every day, but it was also challenging enough to tax listeners' attention abilities. To explore whether suppression is sensitive to object-based statistical regularities in auditory environments, three distractor conditions were implemented. In the "High" condition, the distracting sound with a higher probability of occurring was present in the scene, in the "Low" condition the low-probability distracting sound was present, and in the "None" condition no distracting sound was present in the scene. This last condition was included

as a control condition to confirm that both distractors disrupted task performance. Highest accuracy and fastest RTs were thus expected in the “None” condition. If participants learned to suppress a distractor that occurred most frequently, then higher accuracy and faster RTs were expected in the “High” condition relative to the “Low” condition. Such a pattern of results would suggest that the auditory attention system is capable of suppressing a salient distracting sound that occurs more regularly, which could be advantageous for remaining focused on task-relevant information.

Experiment 1

Method

Participants. Twenty-four participants were recruited from introductory psychology courses at The Ohio State University and received partial course credit as compensation for their time. Data from two participants were excluded from analyses: one due to being a non-native speaker of English, and one due to misunderstanding task instructions. As a result, analyses were restricted to data from a total of 22 participants (11 male). All participants self-reported normal hearing abilities.

Stimuli. Speech stimuli consisted of the spoken numbers one through nine, excluding five. Each number was spoken three times by three different males and three different females. The environmental (distracting) sound stimuli were obtained from online recordings and consisted of a guinea pig squeak and a bird tweet. These sounds were selected to minimize overlap with the frequencies of human speech, and most of their energy was concentrated above 2000 Hz.

All sounds were time stretched in Adobe Audition CC (2017) to have a duration of 300 *ms*. Sounds were then lateralized at one of three angles (0, 90, 180) using HRTFs from Kayser, Ewert, Anemüller, Rohdenburg, Hohmann, and Kollmeier (2009). These lateralized versions of each sound were then normalized to 68 dB using a custom script written in Praat (Boersma, 2001). Finally, the perceptual loudness of all lateralized versions of the two environmental sounds were manually equated to match the speech stimuli.

A custom Python script was used to create the complex scenes. First, individual sound files were combined to form streams for each talker or sound type. Each stream consisted of 20

sound files from the same location with 50 *ms* of silence between each sound. For the environmental sounds, the same sound was repeated 20 times. For the speech stimuli, numbers were chosen such that only even or odd numbers were presented in each stream, and the same number was never presented twice in a row. Finally, individual streams were combined to form the full four-sound scenes.

There were three conditions in this experiment that determined scene creation. In the "None" condition, no distractor was present, so scenes consisted of four talkers. Because the primary task was to find the talker that was a different gender (target voice), three voices were the same gender, and one voice was a different gender. Following Wang and Theeuwes (2018), in the "High" condition, one of the distractors occurred with a high probability (70%), and in the "Low" condition, the other distractor occurred with lower probability (30%). In these two conditions, scenes consisted of the appropriate distracting sound, two voices of one gender, and one voice of the other gender.

Multiple constraints were placed on scene formation to ensure there were no confounds that could influence performance. Each of the six voices was designated as the target voice approximately the same number of times throughout the experiment. When each voice was the target voice, the correct response was "even" on 50% of trials and "odd" on the other trials. When there were four voices in the scene (None condition), two voices spoke even numbers, and the other two spoke odd numbers. When there were only three voices and one distractor in the scene, one of the nontarget voices was assigned to speak the same category of numbers as the target. This was done to encourage participants to attend to the voices rather than to the content of the speech.

Three locations were used in each scene (left, right, center), but four sounds were required in order for participants to perform the task. As a result, one location always contained two sounds. This not only permitted clear perception of three distinct locations via headphones, but it also enabled us to examine how performance changed depending on whether the distracting sounds were presented in isolation or paired with one of the voices. Several rules were established to balance the influence of this location pairing. The target voice was never permitted to occur in the same location as a distracting sound. It was presented in its own location on 50% of trials, and with one of the nontarget voices on the other 50% of trials. Nontarget voices were permitted to occur in the same location as one of the distractors, and did so randomly. Each of the three locations contained two sounds about the same number of times throughout the experiment. The target voice was presented in each of the three locations the same number of times throughout the experiment. The high-probability and low-probability distractors occurred in each location the same number of times.

Because all sounds within a stream were the same duration and separated by the same interstimulus interval, the onsets of three of the streams in each scene were staggered. One of the nontarget voices was always the first stream to play, and the remaining streams were randomly assigned delays of 100, 225, and 380 *ms*.

Procedure. The experiment took place in a sound-attenuated room and all stimuli were presented via Sony MDR-V900 dynamic stereo headphones. Participants were first introduced to the male and female voices by listening to each talker say each number while seeing the words "Male" or "Female" and a corresponding cartoon image on the screen. After a brief familiarization session with the individual number stimuli, participants listened to one talker at a

time say a sequence of three numbers. They were asked to categorize each voice as “Male” or “Female” by clicking the appropriate button on the screen. Feedback was provided for incorrect responses. Participants had to correctly categorize the voices on 15/18 trials in order for their data to be included in the final analysis.

Participants then completed 20 practice trials consisting of full scenes. The first 15 practice trials were “None” trials, and did not contain any distracting sounds. The last five trials consisted of three voices and one distracting sound (a blender). This sound was chosen to be different to prevent participants from having additional exposure to either of the distractors. Participants were asked to listen to each scene and find the voice that was a different gender. Once they found that voice, they were asked to press “o” on the keyboard if the voice was saying odd numbers, or “e” if it was saying even numbers. If participants responded incorrectly, the word “Incorrect” appeared in the center of the screen in red font for 500 *ms*. After each response, the onset of the next trial was delayed by a variable intertrial interval ranging from 50 *ms* to 1250 *ms*.

During the practice trials, participants were also introduced to the visual feedback that was present on every trial. In an effort to reduce speed-accuracy trade-offs, participants were encouraged to focus both on speed and on accuracy. Feedback was designed to encourage participants to keep their mean accuracy at or above 75%, and mean response times at or below 4500 *ms*. These values were chosen on the basis of participant averages during pilot testing. Two ovals were presented in the bottom right-hand corner of the screen: one labeled “Accuracy,” and one labeled “Speed.” Participants were told that their goal was to make sure both ovals remained bright green. If either average accuracy or response times fell below the goal averages, the

corresponding oval would begin to transition from bright green to yellow-green, yellow, orange, and finally to red. The running averages for accuracy and response times were used to determine color transitions. Once average performance improved, the appropriate oval would begin to transition back toward bright green.

After the practice trials, participants began the main experiment. The high/low probability assignments for each distractor were counterbalanced across participants. The experiment was broken into three blocks of trials so participants could take regular rest breaks. Three "burn-in" trials were included at the beginning of each block to allow participants to reorient to experimental conditions. Within each block there were 12 trials in the None condition, 12 trials in the Low condition, and 28 trials in the High condition for a total of 156 trials. Trials were pseudo-randomly-distributed within each block.

Results and Discussion

Experiment 1 was designed to determine if suppression can capitalize on object-based regularities in complex auditory scenes to improve selective attention. If the suppression mechanism learned to attenuate the most frequently-occurring distractor sound, higher accuracy and faster RTs were expected in the High condition relative to the Low condition. All analyses were completed using JASP (JASP Team, 2018). Trials on which participants failed to respond before the sound stopped playing were excluded from analyses. Responses occurring within the first 300 *ms* and intra-individual RT outliers (± 3 *SDs*) were also excluded from analyses. Differences between conditions were assessed using Bayesian analysis of variance, and Bayes factors (BF_{10}) are reported to quantify the evidence in favor of the predicted outcome. According to popular benchmarks, a BF below 1 indicates evidence in favor of the null hypothesis. A BF

between 1 and 3 indicates minimal evidence in favor of the alternative hypothesis, 3-10 indicates moderate evidence, 10-30 indicates strong evidence, 30-100 indicates very strong evidence, and >100 indicates extreme evidence in favor of the alternative hypothesis (Lee & Wagenmakers, 2013).

Overall, accuracy data were uninformative as differences due to condition were minimal, $BF_{10} = 0.88$. Accuracy was above 85% in all conditions, and differed by no more than 4% across conditions. There was no meaningful difference between the High and Low conditions ($BF_{10} = 0.28$), indicating that the statistical regularities did not influence accuracy. This lack of differences can likely be attributed to highly accurate performance across conditions.

Unlike the accuracy results, RT data suggest that the statistical regularities helped reduce the degree of distraction. As can be seen in panel A of Figure 1, there were extremely large differences in RT due to condition, $BF_{10} = 451.08$. RTs were slower when there was a distractor present in the scene, suggesting that the addition of a distracting environmental sound made it more challenging for listeners to focus on task-relevant information. RTs were 348 *ms* slower in the Low condition relative to the None condition, $BF_{10} = 132.97$, and 174 *ms* slower in the High condition relative to the None condition, $BF_{10} = 1.98$. Of greater interest, RTs in the Low condition were substantially slower (174 *ms*) relative to RTs in the High condition, $BF_{10} = 17.45$, indicating that the frequency with which the high-probability distractor occurred mitigated its distracting influence. Participants were able to learn that the sound contained no useful information and could then ignore it in order to facilitate task performance.

To assess the validity of these results, an exact replication (N=23) was conducted and the data are presented in panel B of Figure 1. Similar to Experiment 1, accuracy was quite high

(above 87%) in all conditions, and did not differ much (no more than 3%) across conditions, BF_{10} = 1.54. RTs were again slower in the Low (288 *ms* difference, BF_{10} = 79.48) and High (137 *ms* difference, BF_{10} = 3.27) conditions relative to the None condition, indicating that the environmental sounds used in this study were distracting and disrupted task performance. The primary result was also replicated, with substantially slower RTs in the Low condition relative to the High condition (151 *ms*, BF_{10} = 13.59). The Bayes factor is slightly smaller than that for the original experiment, but still indicates strong evidence in favor of a meaningful RT difference between the High and Low conditions, and confirms that the statistical regularities helped reduce distraction by the high-probability distractor in this task.

These results demonstrate that the suppression mechanism is not only useful in simple listening situations (e.g. Allen et al., 2011; Klein & Stolz, 2015), but that it also aids task performance in complex listening situations that demand sustained attention. The auditory attention system seems sensitive to object-based statistical regularities, and can use those regularities to suppress frequently-occurring distracting sounds. Given the consistency in the results from Experiment 1 and its replication, it appears that these results are stable and suggests that this experimental paradigm is useful for investigating suppression in auditory selective attention.

Experiment 2

Experiment 1 provided evidence suggesting that participants can suppress a salient distracting sound that frequently appears in a complex listening environment. This result not only parallels the visual results of Wang & Theeuwes (2018), but it also among the first demonstrations of suppression in auditory selective attention in a multi-object listening situation. Although this was a step towards approximating a natural listening environment, stimuli were presented via headphones, which is not how listeners typically experience complex auditory scenes. Sounds usually come from distinct sources and locations, which is known to influence stream segregation (McDonald & Alain, 2005). If streams are successfully segregated, then attention can more easily be focused on a particular sound source. The goal of Experiment 2 was to evaluate whether the pattern of results from Experiment 1 remained the same for a free-field listening environment using loudspeakers. As a result, predictions were identical to those for Experiment 1, with faster RTs in the High condition relative to the Low condition being indicative of suppression.

Method

Participants. Twenty participants (14 male) were recruited from the same population described in Experiment 1.

Stimuli. Stimuli were identical to those for Experiment 1, except that non-lateralized (mono) versions of the individual sound files were used.

Procedure. The experiment took place in a sound-attenuated room and all stimuli were presented via three JBL 305P MKII 5” powered studio monitors. The speakers were placed at ear-level in a semicircle approximately 75cm away from the participant’s head. One speaker was

located directly in front of the other participant, and the other two speakers were located 90° to the left and to the right of that central speaker, such that they were 180° apart from each other. Each of the mono files was presented from a different speaker, with the exception of the location that contained two sounds on each trial. All other procedures were identical to those for Experiment 1.

Results and Discussion

The purpose of Experiment 2 was to see if the results of Experiment 1 would replicate when the stimuli were presented via loudspeakers instead of headphones. Similar to Experiment 1, differences in accuracy were again minimal, $BF_{10} = 0.19$. Accuracy was near 90% in all conditions, and did not differ by more than 2% across conditions.

RT data also parallel the findings of Experiment 1 and suggest that the benefits of having statistical regularities in complex scenes generalize to free-field listening environments. As can be seen in panel C of Figure 1, RTs were slower overall, but there were still large differences in RT due to condition, $BF_{10} = 2403.45$. Of greater importance, RTs were slower in the Low condition relative to the High condition (112 *ms* difference, $BF_{10} = 3.93$), providing reasonable support of the claim that participants suppressed the distracting sound that occurred more frequently. Although this effect generalized to a free-field listening environment, the statistical evidence accompanying the effect was weaker in Experiment 2 relative to Experiment 1, likely due to genuinely smaller effects in Experiment 2. Most participants (16/20) exhibited longer RTs in the Low condition relative to the High condition, but the magnitude of this difference was considerably smaller in Experiment 2. In Experiment 1 the median RT difference between the High and Low conditions was 195 *ms*, whereas in Experiment 2 the median difference was 139

ms. The Bayes factor still provides moderate evidence in favor of a meaningful difference between the High and Low conditions, but the current results suggest that object-based statistical regularities may not be as useful in free-field listening environments where spatial cues are stronger.

General Discussion

The present investigation provides another behavioral demonstration of suppression in auditory selective attention, and expands on earlier results to show that suppression is sensitive to object-based statistical regularities in extended, complex listening environments. The listening task that was developed for this study required participants to search multi-object scenes for a target voice while simultaneously ignoring a distracting environmental sound. This task not only approximated the complexity of everyday listening situations, but its challenging nature also brought out differences due to statistical regularities. If the task were too simple, it may have been easy for participants to ignore the distractors, and differences between conditions may not have emerged. In all three datasets, it took participants longer to find the target voice when a less-predictable distracting sound appeared in the scene relative to when a highly-predictable distractor was present. These results suggest that the auditory attention system learned that the high-probability distractor contained no useful information, and then suppressed it to facilitate attention to task-relevant stimuli.

Despite this clear evidence in favor of suppression of the high-probability distractor, it is useful to explore how the suppression mechanism evolved and responded to the statistical regularities over the course of the experiment. The higher regularity of the high-probability distractor should lead to it being learned and suppressed first, but participants also gained experience with the low-probability distractor throughout the course of the experiment. Because the low-probability distractor also reliably contained no useful information, it would benefit the attention system to suppress it as well. Figure 2 presents RTs in each condition across time (by block), and shows that the difference between the High and Low conditions in Experiment 1

sharply decreased from the first block (266 *ms*) to the second (30 *ms*). Similar improvements occurred in the replication experiment and in Experiment 2. These data indicate that participants quickly learned to suppress the high-probability distractor within the first block of trials, then subsequently suppressed the low-probability distractor. This pattern of results supports the hypothesis that suppression in auditory selective attention is sensitive to object-based statistical regularities, and suggests that participants suppressed the highly-probable distracting sound more quickly and to a greater degree relative to the less-probable distractor.

Even though the difference between the High and Low conditions became smaller over the course of the experiment, RTs continued to be faster when the high-probability distractor was in the scene. This suggests that auditory suppression can operate on multiple objects, but suppression of the more predictable distractor will be prioritized. Because selectively attending in complex listening environments is a cognitively-demanding task, prioritizing suppression of more predictable distracting signals is likely an efficient strategy. Overall, these results not only suggest that suppression is an active mechanism when attention is sustained for longer periods of time, but also support the idea that suppression in auditory selective attention is sensitive to object-based statistical regularities in the environment.

Although suppression has not been studied extensively in audition, a number of studies have proposed suppression as a crucial mechanism underlying auditory attention. For example, Alain and Bernstein (2008) proposed a theory of attention stating that attention increases stream segregation via an enhancement mechanism and a suppression mechanism. They claimed that task-relevant acoustic information is prioritized at the expense of task-irrelevant information. The current study provides data to support the suppression component of that theory, in addition

to suggesting a manner in which suppression can occur. Participants knew that the distracting sounds were irrelevant information, and learned to suppress the more frequent distractor in order to facilitate focusing on the task-relevant information. Even in environments where the relevant and irrelevant information are not always immediately known, experience with a scene can make those distinctions clear. Imagine a situation where you are sitting in a meeting trying to focus on what someone is saying, but there is construction noise outside of the room. At first the noises might be extremely distracting, but if the noises remain the same after several minutes, it generally becomes easier to tune them out. This phenomenon could result from a suppression mechanism using statistical regularities in the environment to attenuate frequently-occurring distracting signals that convey no useful information.

Suppression has also been proposed as a potential mechanism underlying the auditory negative priming effect, which refers to impaired performance in response to previously-irrelevant stimuli. According to a review by Frings, Schneider, and Moeller (2014), there are two primary accounts that explain this effect: an inhibition-based account and a retrieval-based account. The inhibition theory assumes the stimulus was actively suppressed by selective attention while it was irrelevant, so the suppressed representation must then be activated when it becomes relevant, leading to negative priming. The retrieval-based account claims that the memory trace of the probe trial contains a "distractor" tag that creates interference when the irrelevant stimulus then becomes relevant on a subsequent trial. Evidence for both accounts has been demonstrated using a variety of tasks, and many researchers agree that both processes are likely involved in negative priming. Although the current investigation was not designed to

probe the mechanisms underlying auditory negative priming, data provide additional support for the inhibition account by demonstrating active suppression of an irrelevant signal.

The primary contribution of the present investigation is evidence that suppression in auditory selective attention is sensitive to the regularities of auditory objects in multi-sound scenes. These data complement earlier findings that suppression seems to be sensitive to spatial regularities in auditory scenes consisting of two voices (Allen et al., 2011). They observed impaired performance when the target syllable came from the expected location of the concurrent masking syllable. Although the scenes were relatively simple and brief, the study still obtained evidence that suppression in auditory selective attention may be sensitive to statistical regularities in distractor location. However, it should be mentioned that Jones and Litovsky (2008) failed to find effects related to distractor location expectations when implementing a similar experimental design using slightly longer stimuli (spondees), so additional research is necessary to determine the nature of sensitivity to statistical regularities in distractor location. The design of the present investigation could easily be adapted to investigate the time course of suppression effects and obtain a clearer understanding of the operation of statistical learning in auditory selective attention.

The pattern of RT results obtained in the current study replicate results from the visual statistical learning experiment of Wang & Theeuwes (2018), suggesting that a common suppression mechanism may underlie the effects observed in both modalities. RTs were approximately 3000 *ms* longer in the current investigation than in the visual study. One of the key differences between vision and audition is that visual objects are often immediately available in a scene, but auditory objects must be formed over time. In fact, Carlyon, Cusack, Foxton, and

Robertson (2004) demonstrated that the buildup of auditory streaming can take as long as 10 seconds. Thus in order to adapt a visual task for audition, longer scenes were used in the current study. However, the fact that the pattern of results was identical suggests that suppression was at work in both situations, and that a common mechanism may be shared across modalities that operate over different time courses. This supports earlier proposals that similar principles can explain both visual and auditory attention (e.g., Shinn-Cunningham, 2008).

Another contribution of the current study is the demonstration of a suppression effect that operates on longer, more complex scenes. Previous demonstrations of suppression have typically implemented simple scenes consisting of the simultaneous presentation of two sound sources. Some of these scenes have consisted of sequences of tones (Bidet-Caulet, Mikyska, & Knight, 2010; Melara et al., 2012), whereas others have included dichotic presentation of two voices each saying a single word (Nolden et al., 2019). While these studies have provided important preliminary data in favor of suppression in auditory selective attention, most real-life situations require listeners to maintain focus on a particular sound source for a longer period of time, and often among a larger variety of distractions. Future investigations of the mechanisms underlying auditory selective attention should continue to generalize the results of earlier studies to more complex listening situations.

Given the stability of the suppression effect demonstrated in all three datasets, a next step is to understand which aspects of the scene drive that effect. Exploratory analyses on the data from Experiment 1 and the replication of that experiment suggest that the RT differences between the High and Low conditions were dependent on the distractor being presented in its own location rather than sharing a location with one of the nontarget voices (see panels A and B

of Figure 3). RTs were substantially slower when the low-probability distractor had a distinct location, suggesting that it may have been pulling attention away from locations that contained task-relevant information (target and nontarget voices). When the low-probability distractor shared a location with a nontarget voice, it likely still grabbed attention, but in that situation attention was drawn toward a location that contained task-relevant information. A similar increase in RT was not observed when the high-probability distractor had its own location, suggesting that suppression prevented the distractor from capturing attention.

As can be seen in panel C of Figure 3, the source of the differences between the High and Low conditions changed when the stimuli were presented over loudspeakers instead of headphones. Rather than an RT cost of the low-probability distractor having its own location, there was an RT benefit when the high-probability distractor shared a location with one of the nontarget voices. Whenever a distractor was paired with a voice, each of the three voices in the scene was presented from a different loudspeaker. This physical separation in a free-field listening environment likely facilitated segregation of the three voices. This in turn improved task performance, but only when the distracting sound was suppressed so as to not mask the voice coming from the same speaker. Although sounds were lateralized to simulate different locations when presented over headphones, it was still a weaker localization cue than actual physical separation in a free-field listening environment, which could explain why the pattern of results changed when stimuli were presented via loudspeakers. However, because the current investigation was not designed to explore these possibilities, the number of observations in each cell was grossly imbalanced and statistical support for the effect was uninformative. Future investigations should draw on distractibility research (e.g., Macken, 2014) to better understand

the situations in which a stimulus causes distraction, and should use that information to make predictions about how suppression can reduce that distractibility.

Another goal for future studies is to determine whether enhancement of auditory objects is similarly sensitive to statistical regularities in the environment, and to understand how suppression and enhancement might interact with one another. Are there certain individual characteristics that determine whether one mechanism will be dominant? Does the listening situation predict which mechanism will be more beneficial? In order to understand the how enhancement and suppression interact to enable selective attention, it is essential to first understand how each mechanism operates in isolation. The present investigation not only provided a novel behavioral technique for isolating the effects of suppression from those of enhancement in auditory selective attention, but it also demonstrated that suppression is sensitive to statistical regularities in auditory scenes that more closely approximate the complexity and challenges of everyday listening.

References

Addleman, D. A., & Jiang, Y. V. (2019). The influence of selection history on auditory spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 45*(4), 474-488.

Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience, 5*, d202-212.

Alain, C., & Bernstein, L. J. (2008). From sounds to meaning: The role of attention during auditory scene analysis. *Current Opinion in Otolaryngology & Head and Neck Surgery, 16*, 485-489.

Allen, K., Alais, D., Shinn-Cunningham, B., & Carlile, S. (2011). Masker location uncertainty reveals evidence for suppression of maskers in two-talker contexts. *The Journal of the Acoustical Society of America, 130*(4), 2043-2053.

Best, V., Ozmeral, E. J., Kopčo, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *PNAS, 105*(35), 13174-13178.

Bidet-Caulet, A., Mikyska, C., & Knight, R. T. (2010). Load effects in auditory selective attention: Evidence for distinct facilitation and inhibition mechanisms. *NeuroImage, 50*, 277-284.

Bressler, S., Masud, S., Bharadwaj, H., & Shinn-Cunningham, B. (2014). Bottom-up influences of voice continuity in focusing selective auditory attention. *Psychological Research, 78*, 349-360.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 115-127.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 25*(5), 975-979.

Frings, C., Schneider, K. K., & Moeller, B. (2014). Auditory distractor processing in sequential selection tasks. *Psychological Research, 78*, 411-422.

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D'Esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of Cognitive Neuroscience, 17*(3), 507-517.

Ihlefeld, A., & Shinn-Cunningham, B. (2008). Disentangling the effects of spatial cues on selection and formation of auditory objects. *The Journal of the Acoustical Society of America, 124*(4), 2224-2235.

Jones, G. L., & Litovsky, R. Y. (2008). Role of masker predictability in the cocktail party problem. *The Journal of the Acoustical Society of America, 124*(6), 3818-3830.

Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., & Kollmeier, B. (2009). Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. *EURASIP Journal on Advances in Signal Processing*, 298605, doi: 10.1155/2009/298605

Klein, M. D., & Stolz, J. A. (2015). Looking and listening: A comparison of intertrial repetition effects in visual and auditory search tasks. *Attention, Perception, and Psychophysics, 77*, 1986-1997.

Koch, I., & Lawo, V. (2014). The flip side of the auditory spatial selection benefit: Larger attentional mixing costs for target selection by ear than by gender in auditory task switching. *Experimental Psychology, 62*(1), 66-74.

Lee, M. D., & Wagenmakers, E.-J. (2013). Bayesian cognitive modeling: A practical course. Cambridge University Press.

Lewald, J., Hanenberg, C., & Getzmann, S. (2016). Brain correlates of the orientation of auditory spatial attention onto speaker location in a "cocktail-party" situation. *Psychophysiology, 53*, 1484-1495.

Macken, B. (2014). Auditory distraction and perceptual organization: Streams of unconscious processing. *PsyCh Journal, 3,* 4-16.

Melara, R. D., Tong, Y., & Rao, A. (2012). Control of working memory: Effects of attention training on target recognition and distractor salience in an auditory selection task. *Brain Research, 1430*, 68-77.

Nolden, S., Ibrahim, C. N., & Koch, I. (2019). Cognitive control in the cocktail party: Preparing selective attention to dichotically presented voices supports distractor suppression. *Attention, Perception, & Psychophysics, 81*, 727-737.

Samson, F., & Johnsrude, I. S. (2016). Effects of a consistent target or masker voice on target speech intelligibility in two- and three-talker mixtures. *The Journal of the Acoustical Society of America, 139*(3), 1037-1046.

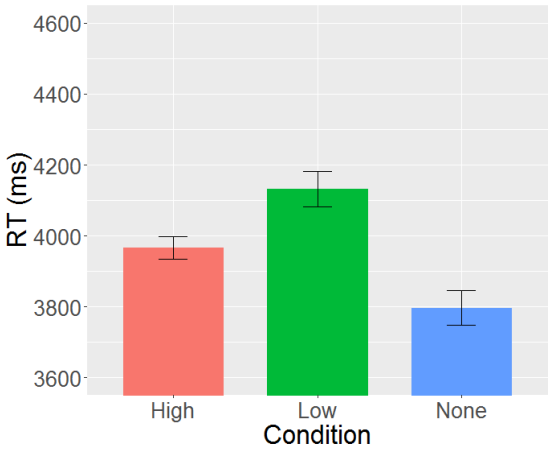
Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12*(5), 182-186.

Wang, B., & Theeuwes, J. (2018). Statistical regularities modulate attentional capture. *Journal of Experimental Psychology: Human Perception & Performance, 44*(1), 13-17.

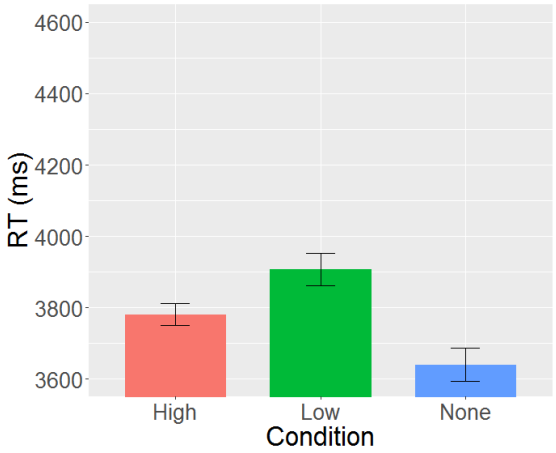
Zimmerman, J. F., Moscovitch, M., & Alain, C. (2016). Attending to auditory memory. *Brain Research, 1640*, 208-221.

Appendix: Figures

A)



B)



C)

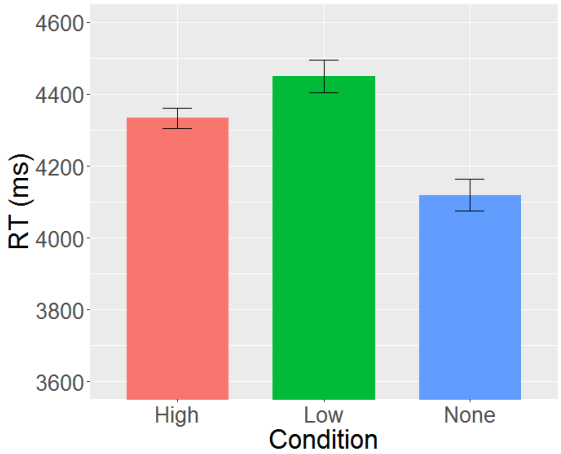


Figure 1. Mean RTs and *SEs* for A) Experiment 1, B) replication of Experiment 1, and C) Experiment 2.

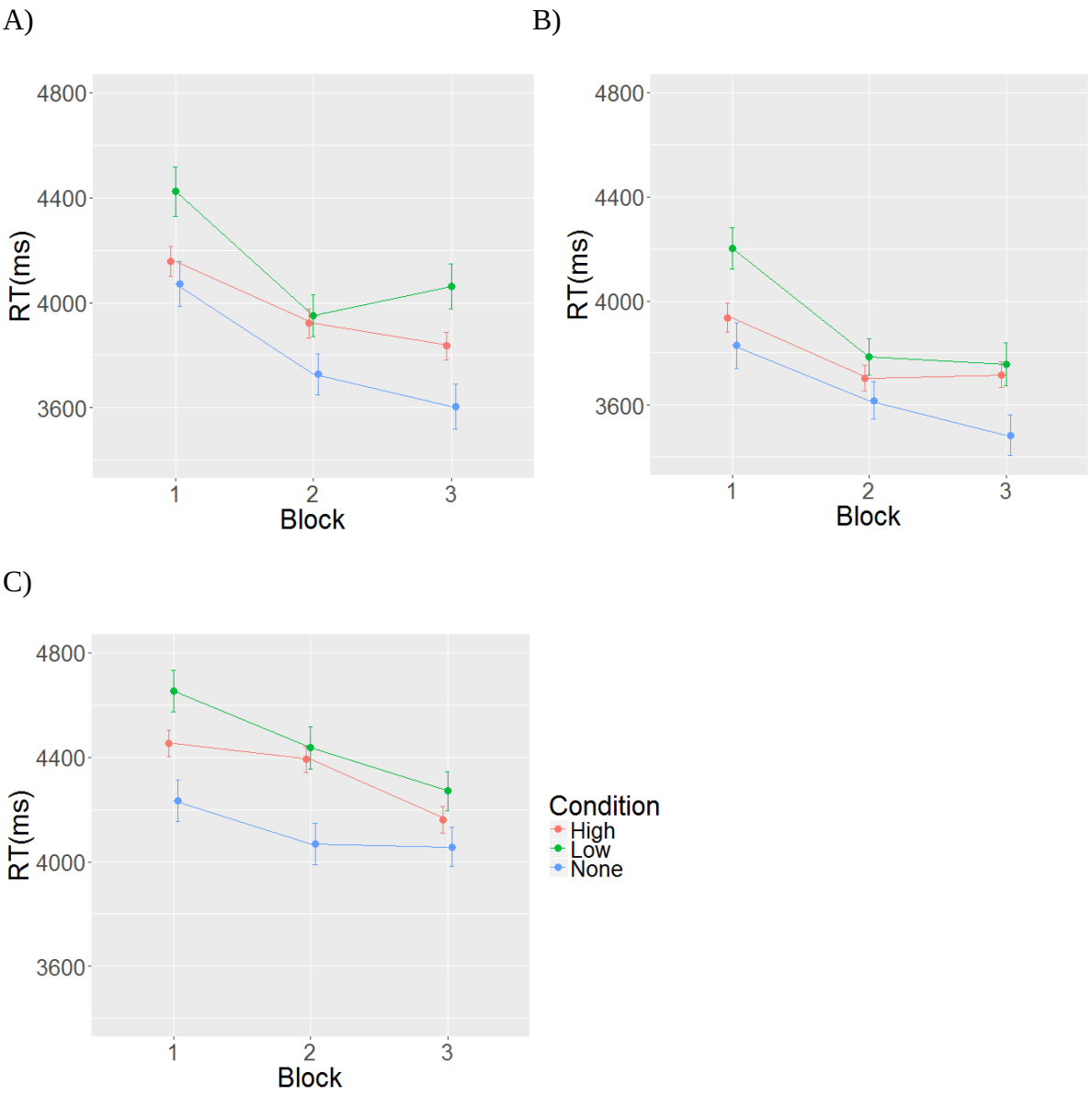
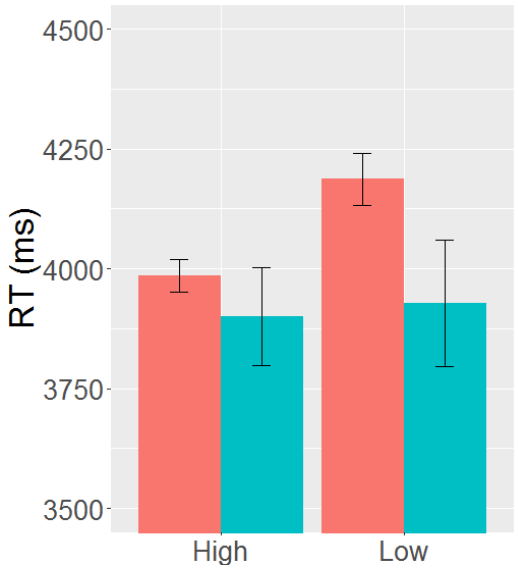
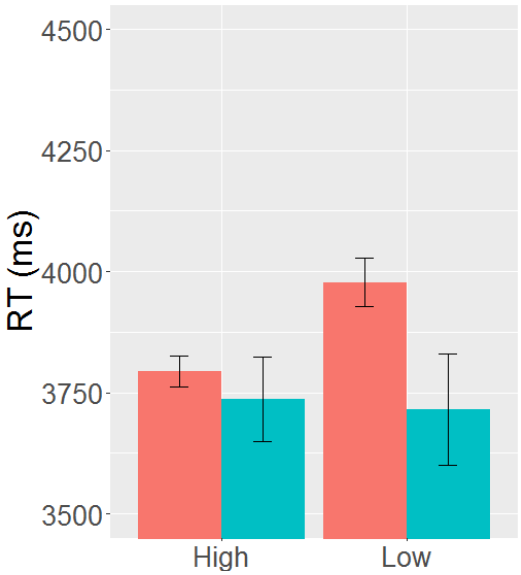


Figure 2. Mean RTs and *SEs* across blocks for A) Experiment 1, B) replication of Experiment 1, and C) Experiment 2.

A)



B)



C)

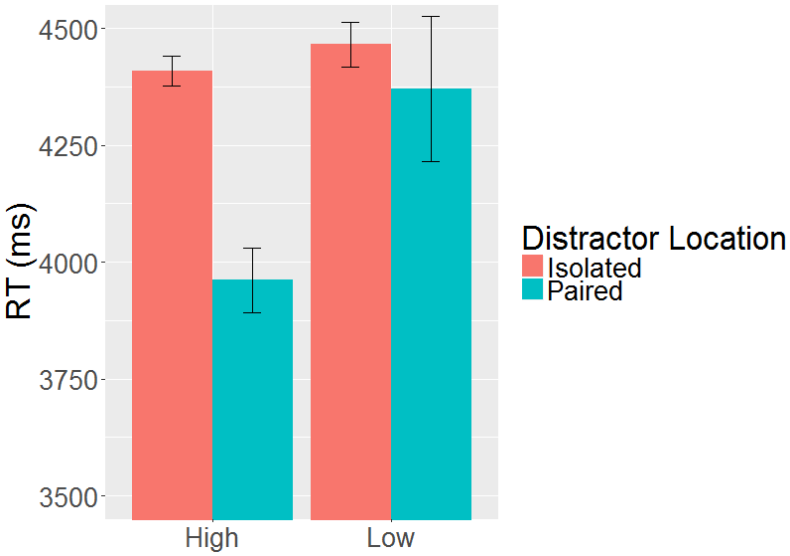


Figure 3. Mean RTs and *SEs* dependent on whether the high- or low-probability distracting sound was presented by itself in a location (Isolated) or whether it shared a location with one of the nontarget voices (Paired).