Modulation of Auditory Capture by the Identity of Visual Priors

Jocelyn Huerta

Mentor: Dr. Joy Geng
Department of Cognitive Science at the University of California, Davis

Abstract

Research has shown that sounds can distract from visual tasks when they occur in a different location from a visual object, but they also facilitate visual search when they are semantically consistent with a visual target object. However, it is not yet known how knowledge of a visual object's attentional priority influences processing of auditory stimuli during visual search. In the present study, we asked subjects to search for a visual target embedded on visual objects (ducks, frogs). We manipulated the attentional priority of these objects by making the target more likely to occur on one of these objects. We hypothesized that sounds that were semantically congruent with the more likely target object (e.g., a quack or ribbit) would facilitate search, despite being completely task-irrelevant and spatially ambiguous. We further predicted that congruent sounds would facilitate saccadic eye movements towards the visual target. Preliminary results are consistent with our hypotheses and suggest that multisensory integration of real-world objects is automatic and attentional priority in one modality influences sensory processing of congruent information in another modality.

Keywords

Visual Search · Visual Stimuli · Auditory Stimuli · Distractors

Introduction

Imagine you are looking for a friend in a coffee shop. You are likely to find her more quickly if you also hear the sound of her voice. Similarly, you are likely to find an object you are looking for (e.g. keys) if you hear a noise in the direction of the object (Iordanescu et al., 2010) . These ordinary examples show the effects of multisensory integration, a sensory from one modality (e.g. vision) influencing the processing of another, on visual search (Koelewijn et al., 2010; Talsma et al., 2010).

In one seminal study, Spence & Driver (1997) looked at the importance of audiovisual interactions on attention by conducting a study that effectively cued the location of visual targets. In the experiment, participants judged the elevation of lights or sounds following a task-irrelevant cue either in the same modality (e.g., light-cue, light-target) or the opposite modality

(e.g., sound-cue, light-target). The setup, as shown in Figure 1, allowed Spence and Diver (1997) to determine if spatially localized sounds influenced target judgments in the visual modality. The results showed that the participants had quicker and more accurate judgments when the sounds were spatially congruent with the target. Although this research is fundamental for understanding crossmodal attention, Spence and Driver (1997) solely used white noise for the auditory targets and LEDs for the visual targets.

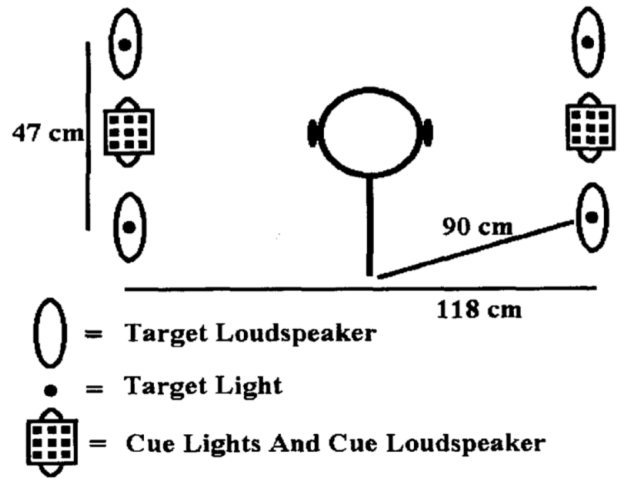

Figure 1. (Spence & Driver,1997) In this study, participants were presented with irrelevant cues on either side followed by the target (auditory or visual). Participants were expected to judge target elevation.

This is an important limitation to address, as some studies have found that multisensory integration can be affected by meaningful content within stimuli. Iordanescu et al. (2010) looked at the effects of characteristic sounds on visual search by using realistic stimuli. In the experiment, participants were presented with the name of the target aurally followed by a search display of objects (e.g. frog, cat, dog) accompanied with a sound that was either characteristic of the target, a distractor sound, or no sound. The results showed that when the target was paired with the characteristic sound, the target was found more quickly. The characteristic sound guided the eye movements towards the target significantly faster, even though sounds were spatially ambiguous and not localized to the visual target. Additional studies have suggested the importance of characteristic sounds in various aspects of visual search (Iordanescu et al., 2008). These studies communicate the importance of characteristic sounds and realistic stimuli (as seen in Figure 2) in visual search.

Figure 2. (Iordanescu et al., 2010) In the study, participants were expected to look for a visual target after they had been presented with the target name aurally (e.g. dog). Participants were instructed to find the target in the search display.

This previous work has been necessary to our understanding of the effects of sounds on visual search, even though little work has been done to bridge the gap between studies that have used localizable sounds and those that have used realistic multisensory objects.The study by Spence and Driver (1997) influenced the localization of sounds in other studies. Koelewijn et al. (2009) based their experiment on the Spence and Driver (1997) paradigm, but concentrated on attentional capture and found exogenous capture of visual attention by irrelevant sounds but used dots and white noise rather than realistic stimuli. Another study by Koelewijn et al. (2009) localized the visual and auditory cues but utilized white noise for the auditory cue and a gray bar for the visual cue. Studies done by Iordanescu et al. (2008), on the other hand, have used realistic stimuli to show the effects of sounds on visual search but did not localize sounds. The present study incorporates the use of realistic stimuli and the localization of sounds.

Realistic stimuli and localizable sounds are both important in order to build models of attention and perception that are more generalizable from the laboratory to the real-world. In a natural environment, you are likely to see an object before you hear it. Just like you are likely to see a friend across the hall before they call out to you. The present study shows a visual stimulus before the sound while using realistic stimuli and spatially localizable sounds. We are interested in investigating how knowledge of object location influences how sounds capture attention. Knowing the location of a distractor object (e.g. dog) may lead to less unwanted attentional capture by the characteristic sound (e.g. “ruff”) from that object’s location because the object was seen and its sound is expected. In the Busse et al. (2005) study, subjects were expected to fixate on a point while directing covert attention to the left or right side of the monitor. The study showed that attentional priority can spread from one modality to another. The present study aims to investigate how the spread of attentional weighting from one modality to another might be modulated by semantic information within realistic stimuli.

We created a task where the Landolt C targets were overlaid on top of the realistic stimuli (ducks, frogs) and manipulated the probability that the target would appear on each visual stimulus in each trial. For instance, the target could be placed on the frog for 80% of the trials, making it the high probability target location, and on the duck for 20% of the trials, making it the low probability target location. On each trial, participants had to respond via keyboard with the orientation of the target Landolt C. On each trial, there was also a task-irrelevant sound (quack, ribbit, white noise) located on the side of the target or the side of the distractor. This resulted in a total of eight conditions in a two (target location: high or low probability) by two (sound: characteristic or white noise) by two (sound location: target or distractor location) design.

We tracked participants' first saccades and predicted that the majority of saccades would be directed towards the high probability animal across all conditions (regardless of the target location, sound location, or sound type). We also predicted the majority of first saccades to be directed towards the low probability animal congruent with white noise as opposed to the low probability animal congruent with the characteristic sound because we expected lower attentional priority from the low probability animal paired with the characteristic sound.

We also measured reaction time and developed four hypotheses. First, we predicted reaction times to be faster for trials with the target appearing on the high probability animal than the low probability animal. Second, we predicted that the reaction time would be slower when the sound was incongruent to the target and faster when the sound was congruent with the target. Third, we predicted that if the target took in the attentional priority of the characteristic sound from the high probability animal then the reaction time would be faster when the low probability animal was paired with its characteristic sound than the low probability animal paired with white noise because the expected distractor would be suppressed. Lastly, we predicted that if the target was on the low probability animal paired with its characteristic sounds then there would be a faster reaction time for the low probability animal with the characteristic sound than the low probability animal with the white noise.

Method

Participants

Ten undergraduate students (8 female) from the University of California, Davis participated in the experiment. They were recruited from SONA for course credit. They had normal or corrected vision and hearing. Participants were informed of the expectations beforehand without being informed of the aim of the experiment.

Apparatus and Design

Participants were seated in a soundproof room in front of a monitor. Eyelink 1000 eye-tracking was used to track saccades. The device was placed under the monitor. The experiment was run on PsychoPy and was navigated from another room. The sounds were presented from lateralized speakers placed to the left and to the right side of the monitor. There were a total of eight conditions that were manipulated through sound, visual objects, and target location. There were two visual objects (duck, frog) accompanied by a sound that was either congruent or incongruent

with the target location. The sound was either characteristic of the visual object (quack, ribbit) or white noise. There were a total of 560 trials. The participant data was analyzed through Python.

Procedure and Stimuli

Participants were presented with a blank screen for 500ms followed by a one-second preview of the visual objects allowing them to identify the locations of each animal, which were randomly placed in each trial. The duck and the frog were chosen to represent realistic stimuli. The preview was expected to help participants develop predictions about the target location for the ensuing trial. They were instructed to fixate on the white cross in the middle of the screen until it disappeared. After, they were expected to identify the direction of the target Landolt C, which would face up or down, as opposed to the distractor Landolt C, which faced right or left. If the gap of the target C was on the top, subjects were instructed to respond with *h* on the keyboard, and if the gap was on the bottom, they were instructed to respond with *b* on the keyboard. The gap was used as a target to enforce participants' attention on the visual object.

The target was placed on top of the duck or the frog and it was presented after the fixation and preview periods along with a coincidental, characteristic sound (e.g. quack, ribbit) or white noise that was either congruent or incongruent to the target location. The white noise was used to control for the characteristic sounds given that it did not have any meaning to the visual stimuli. The target appeared on the high probability target location 80% of the time, making the visual object the high probability animal, and appeared 20% of the time on the low probability target location making the visual object the low probability animal. The high probability animal was counterbalanced across participants to increase validity. While participants were completing the behavioral task, they were also being observed for their first saccades in each trial.

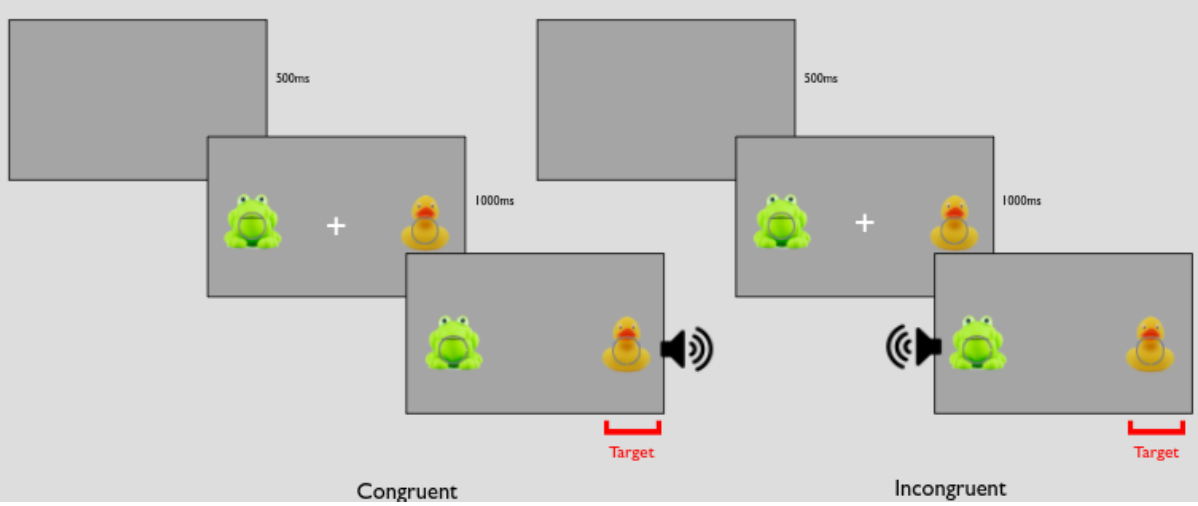

Figure 3. Representation of the present study.
Participants looked at a blank screen for 500ms, then they were presented with the visual stimuli without the target for 1000ms. They were finally presented with the visual stimuli, the target, and the auditory

Results

Eye-tracking conditions

There was a greater quantity of first saccades towards the high probability animal in contrast with the low probability animal (as seen in Figure 4). In the high probability target location condition, when the sound location was target-congruent and the sound was white noise, there was a higher average percentage of first saccades made toward the target (*M* = 75.4%) than towards the distractor (*M* = 24.5%). When the target was congruent with the characteristic sound, there was a higher percentage of first saccades towards the target (*M* = 73.9%) than the distractor (*M* = 26%). When the target was incongruent with the sound (white noise) there was still a greater percentage of first saccades directed towards the target (*M* = 62.5%) as opposed to the distractor (*M* = 37.4). When the target was incongruent with the characteristic sound there was a greater percentage of first saccades for the target (*M* = 60.4%) than the distractor (*M* = 39.5%).

In the low probability target location condition, there was a higher average percentage of first saccades directed towards the distractor (*M* = 77.4%) than the target (*M* = 22.5%) when the target was incongruent with the characteristic sound. When the target was incongruent with the white noise, there was a greater percentage of first saccades for the distractor (*M* = 76%) than the target (*M* = 23.9%). When the target was congruent with the sound (white noise) there was a greater percentage for the distractor (*M* = 61.8%) than the target (*M* = 38.1%). When the target was congruent with the characteristic sound there was a greater percentage of first saccades for the distractor (*M* = 63.3%) than the target (*M* = 36.6%).

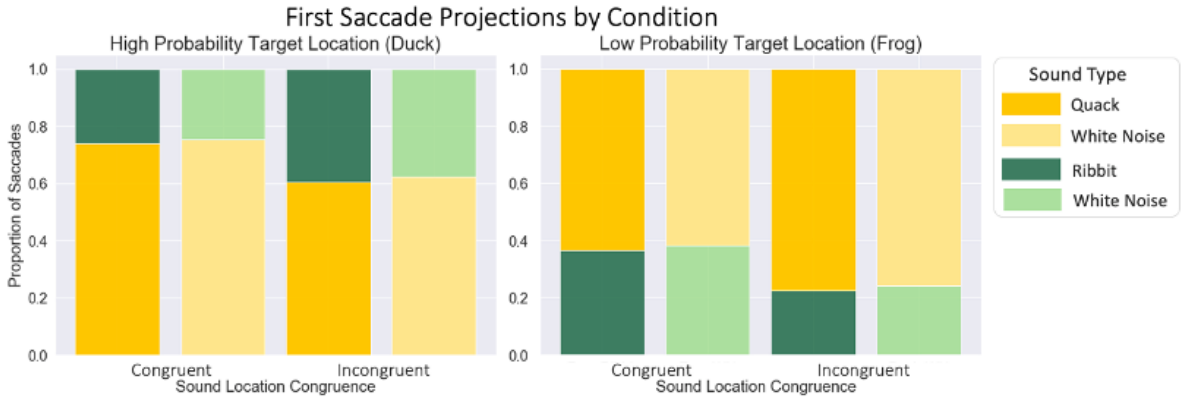

Figure 4. Graph with results on eye-tracking data.
The graph shows the quantity of first saccades and the different conditions for each visual stimuli. The bottom halves of each bar show that the visual and auditory stimuli can be congruent or incongruent. The target visual stimuli is depicted at the bottom half of each graph. In regards to auditory stimuli, the darker bars represent the characteristic sounds while the lighter bars represent white noise.

Reaction time conditions

Overall the reaction times were fastest in the high probability conditions (see Figure 5) and slower in the low probability conditions (Figure 6). The reaction times were fastest when the high probability animal was congruent with the characteristic sound (*M* = 909 ms*)* or when the high probability animal was congruent with the white noise (*M* = 913 ms*).* The reaction times were slower when the high probability animal was incongruent with the characteristic sound (*M* = 1014 ms) or incongruent with the white noise (*M* = 992 ms). In regards to the low probability animal conditions, there were quicker reaction times when the low probability animal was congruent with the characteristic sound (*M* = 1136 ms*)* or congruent with white noise (*M* = 1190 ms). Reaction times were slower when the low probability animal was incongruent with the characteristic sound (*M* = 1222 ms) or incongruent with the white noise (*M* = 1214 ms).

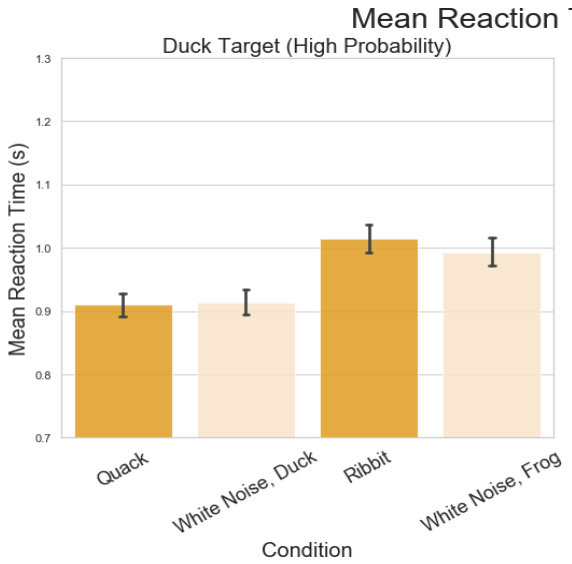

Figure 5. Reaction time for high probability animal. The graph shows the mean reaction time for each condition.

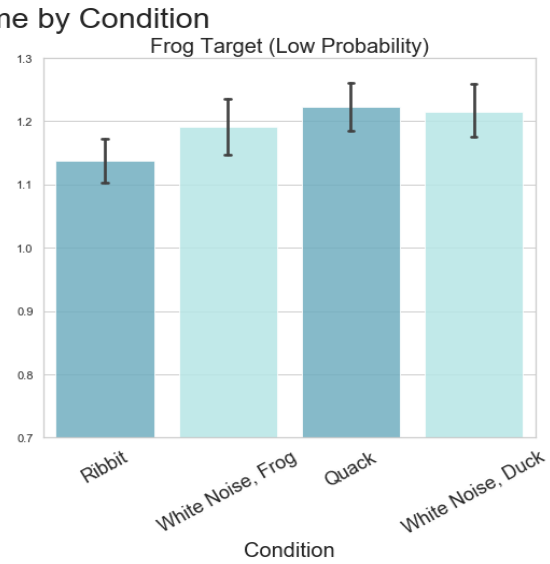

Figure 6. Reaction time for low probability animal. The graph shows the mean reaction time for each condition.

Discussion

We investigated the effects visual priors had on auditory capture. The eye-tracking results suggest that the majority of first saccades were directed to the high probability animal and we were able to determine that participants used the high probability animal as their guide since the target appeared on the high probability target location 80% of the time. Reaction time also proved to be faster when the target was on the high probability animal across all conditions since the target is expected to be on the high probability target location. The results also suggested that auditory and visual congruence had an effect even though participants were instructed to ignore the sounds. Participants' visual search was fastest when the target was spatially and semantically congruent with the auditory stimulus . We did not find that the sound took in the attentional priority of the high probability animal meaning that participants were not able to suppress the low probability animal with its characteristic sound. The results also indicated that there was a

faster reaction time when the low probability animal was paired with its characteristic sound as opposed to white noise. The eye-tracking data indicated that participants looked at the high probability animal first and then quickly looked away resulting in a faster reaction time.

The 80% and 20% probability design was similar to that of Koelewijn et al. (2009), where the visual cue was valid 80% of the time. The probability setup determined whether the low probability animal became the distractor, and determined whether the high probability animal developed attentional priority. The present study included the use of realistic stimuli (frogs, ducks) paired with localized sounds (quack, ribbit, white noise) to represent a real-world environment. The study also looked at sound congruency given there have not been any consistent conclusions on the topic (Tang et al., 2016).

The present study addresses an important gap in the literature on the interactions between the auditory capture of visual attention and the influences of characteristic sounds on visual search. Overall, our data replicated previous research showing that sounds do capture visual attention, even when they are task-irrelevant, and that characteristic sounds also guide attention. However, we expected to see that the presence of visual priors would allow participants to use the visual information to suppress the auditory capture of attention when the sound comes from an object that is less likely to hold the target, though this was not the case in our preliminary data. In real-world situations, this may mean that even when we know the location of a distracting object, a sound that spontaneously comes from that object can still capture visual attention. However, data collection was cut short due to the impact of COVID-19, and more data is needed to draw definite conclusions.

Acknowledgments

I would like to thank the Ronald E. McNair Scholars Program for their support, the Child Family Fund, and the Center for Mind and Brain for funding the project. I also want to give a special thank you to my mentors Dr. Geng and Shea Duarte for guiding and supporting my research experiences.

References

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(51), 18751–18756. https://doi.org/10.1073/pnas.0507704102

Iordanescu, L., Grabowecky, M., Franconeri, S., Theeuwes, J., & Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Attention, Perception & Psychophysics*, *72*(7), 1736–1741. https://doi.org/10.3758/APP.72.7.1736

Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review*, *15*(3), 548–554. https://doi.org/10.3758/PBR.15.3.548

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, *134*(3), 372–384. https://doi.org/10.1016/j.actpsy.2010.03.010

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2009). Auditory and visual capture during focused visual attention. *Journal of experimental psychology. Human perception and performance*, *35*(5), 1303–1315. https://doi.org/10.1037/a0013901

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2009). Competition between auditory and visual spatial cues during visual task performance. *Experimental brain research*, *195*(4), 593–602. https://doi.org/10.1007/s00221-009-1829-y

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, *59*(1), 1–22. https://doi.org/10.3758/BF03206843

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in cognitive sciences*, *14*(9), 400–410. https://doi.org/10.1016/j.tics.2010.06.008

Tang, X., Wu, J., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. Neuroscience & Biobehavioral Reviews, 61, 208–224. https://doi.org/10.1016/j.neubiorev.2015.11.002