Chapter 3

Bias in a Spatial Auditory Attention Task

Cognitive architectures provide a framework for developing models of users interacting with everything from mobile phones [38] to interactive tutors [30]. However, much of the research has been focused on modeling aspects of human cognition associated with using traditional computer interfaces. This includes identifying items on a screen using models of visual attention and perception or modeling the motor skills required for mouse and key presses. Much less work has been done to integrate other cognitive functions, such as spatial auditory attention, which affects how quickly and accurately we attend to the sounds around us. There are many situations where it is useful to simulate auditory attention. For example, hospital emergency rooms make use of auditory alarms to convey important information, and it is helpful to understand when these alarms will be heard, and when they will go ignored. Behavioral experiments have shown that response times to spatial sounds are dependent on the spatial location of the sound [39]. This attentional bias can be modeled as a combination of top-down, or goal-driven processes and bottom-up, or salient, processes. The following chapter describes a new approach for modeling spatial auditory attention

using the AI framework of constraint satisfaction problems and shows how it can be incorporated into the ACT-R cognitive architecture [18]. The work presented in this chapter was originally published in the *Postproceedings of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures* [18], and presented at the *2017 Workshop on Cognition and AI for Human-Centred Design* [40] and the *4th International Workshop on Artificial Intelligence and Cognition* [41]. [1]

3.1 Introduction

Visual attention has attracted a substantial amount of interest from the research community [42, 43, 44], but less work has been devoted to modeling spatial auditory attention [45]. Audition is unique from other senses in that it monitors the environment for sounds happening all around us. The auditory system is particularly useful as an early warning system that can orient attention to things far away and out of sight. This makes it ideal for conveying information about everything from fire alarms to text message notifications. However, this often requires finding a balance between attending to potential threats or opportunities and focusing on a current task.

Computational methods are used to study many aspects of human cognition and behavior, including attention [46, 47, 27]. The goal of this chapter is to better understand spatial auditory attention, particularly examining the attentional processes that govern shifting attention to infrequent distractors. We examine two computational approaches to modeling auditory attention. First, we present a novel method that uses constraint satisfaction problems to model attentional bias as a spatial gradient that results from balancing attention between current goals and important events in the environment. Next, we show how a drift diffusion model captures the ways in which attention develops over time and how this relates to the biases that are modeled in the constraint model.

[1] Code relating to this project can be found at: https://github.com/jaelle/cmsaa

We foresee that this work will advance understanding of basic issues in attention, such as top-down and bottom-up interactions, vigilance, and capacity limitations. Moving beyond the study of spatial auditory attention in isolation, we incorporate it into the cognitive architecture, ACT-R. We substantially extend ACT-R's ability to model auditory attention to include support for spatial auditory attention tasks. This allows us to examine the computational models of spatial auditory attention in the context of the mind as a whole. In our new and sophisticated ACT-R audio module, we create a cognitive agent that simulates human behavior in a spatial auditory attention task, enabling realistic predictions on how fast humans react to sounds. This work opens up new possibilities in designing and optimizing systems for humans where audition is important. For example, there are safety issues when pilots miss critical alarms [48], or when clinicians are unable to distinguish between auditory alarms in a hospital environment [49]. By examining spatial auditory attention in these contexts, it may be possible to predict when such warnings might be ignored.

In Section 3.2, we review the literature related to spatial auditory attention, and cover the fundamentals underlying the computational methods we developed, including constraints, drift diffusion methods, and cognitive architectures. The experiment is described in Section 3.3.1 and the model design is described in Section 3.4. Section 3.5 present the process of fitting the constraint model and drift diffusion model to the behavioral data we collected, as well as a discussion of the results. Section 3.7 describes the development of an extension to the ACT-R audio module and demonstrates how the extension enables ACT-R to accurately simulate human behavior in the spatial auditory attention task.

3.2 Background

We start by providing a brief background on psychology literature related to spatial attention and its computational models. We also give some fundamental information concerning the computational methods we adopt to model spatial auditory attention.

3.2.1 Spatial Attention

Differentiating Top-down and Bottom-up Attention

Almost all attention models distinguish attention that is directed by personal choice from attention that is directed to an event by virtue of it having a salient property, such as a loud sound [50]. This dichotomy is intuitive and has many names in the literature (e.g. top-down/bottom-up, endogenous/exogenous, controlled/automatic [51]. In this chapter, we use the terms "top-down" and "bottom-up". Top-down control regulates information flow based on the current situation and goals in short-term memory by generating a task set to bias processing towards information useful for goal attainment. Bottom-up refers to attention capture that is not guided by current top-down goals (i.e., sounds in the environment that are not actively being listened to). Although the top-down and bottom-up distinction is meaningful, as a practical matter, they are highly interactive [52]. The difficulty of cleanly separating the two processes motivates us to use a computational model, which can examine top-down and bottom-up functions in isolation.

Gradients of Attention

Past work has considered both auditory spatial attention and visual attention at a cognitive level of analysis. Mangun and Hillyard [42] shows that attention can be expressed as a spatial gradient relative to an attended location. Gradients are presumably a byproduct of limited perceptual input capacity, although limitations in

behavioral output may also be relevant [43]. The spatial range of attentional processing is variable [53], and can be modified by directly cuing different size areas [44], or manipulating perceptual or memory loads [54]. Several examples in the literature have shown that auditory spatial cuing decreases reaction times to subsequent targets at a cued location relative to uncued locations [55, 45, 56]. Both Robert J. Zatorre et al. [45] and Rorden and Driver [56] found that target reaction times increased monotonically with greater distance between the cued and target locations. Visual studies suggest that gradients may have a more complex shape ("Mexican-hat"), with reaction times increasing and then decreasing when responding to sounds further from it [57, 58]. Our results show a similar "Mexican-hat" shape in an auditory attention task, but with a much larger spatial range.

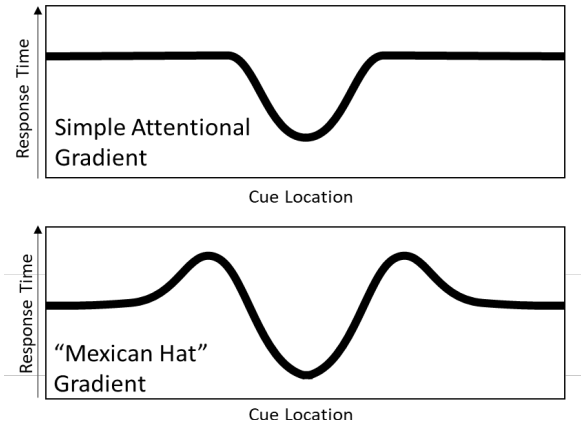

Figure 3.1: Example of attentional gradients proposed in the literature. Some literature has shown that response times at attended locations are faster than those further from the attended location [54]. Other literature has shown evidence for a more complex "Mexican Hat" shape, where response times increase at uncued locations near the attended location and then decrease the further the location is from the attended location [57]

3.2.2 Computational Models of Cognition

Computational models of cognitive processes are beneficial because they require an explicit theory, can reveal hidden assumptions or logical inconsistencies, and simulations can establish proof-of-principle much faster than pilot experiments [6]. Using a

constraint model, we are able to rapidly test ideas of top-down and bottom-up attentional control from prominent verbal models [59], showing how the observed behavior can emerge from top-down and bottom-up interactions. We also use a drift diffusion model of the behavioral task to show how attentional processes that develop over time.

Constraints and Cognitive Models

Constraint programming [16] is a powerful paradigm for modeling and solving combinatorial search problems currently applied with success to many domains, such as scheduling, planning, vehicle routing, configuration, networks, and bioinformatics. The basic idea in constraint programming is that the user states the constraints, and a general-purpose constraint solver is used to solve them. Constraint solvers take a real-world problem, represented in terms of decision variables and constraints, and find an assignment to all the variables that satisfy the constraints. Constraints concern subsets of variables and define which simultaneous assignments to those variables are allowed. For example, in our auditory task, the variables represent the spatial range of possible cue locations, while the constraints limit the amount of attentional bias allocated by top-down and bottom-up attention at each location.

Solutions are found by searching the solution space either systematically, as with *backtracking* algorithms, or use forms of local search which may be incomplete, that is, there is no guarantee they will return a solution. Systematic methods often interleave search and inference, where inference consists of propagating the information contained in one constraint to other constraints via shared variables. The rich variety of finely-tuned algorithms available for constraint problems has made the effort of translating real-world problems into this framework an efficient solving approach.

Constraints have been used before in the context of human cognition to model skilled behavior [10] and learning [11]. An implementation of ACT-R based on con-

straint handling rules, which are closely related to constraints, has been proposed in [60].

In Section 3.4.1, we show how we used this well-established framework to rapidly test various hypotheses about how top-down and bottom-up attention combine to generate observed behavior.

Drift Diffusion Models

To supplement the constraint model of overall attentional bias, we used a diffusion model to explore how cognitive processes develop over time and result in either successful perception, or making an error. The drift diffusion model is used to model the accumulation of information in two-choice tasks as a speed and accuracy trade-off [17]. By comparing the results of our drift diffusion model, with those of our constraint model, we are able to show that the drift rate parameter and boundary separation are predictive of the attentional bias predicted by the constraint model. We also explore how the drift diffusion model can be incorporated into the ACT-R audio module to model individual differences in our behavioral task.

Visual and Auditory Attention in ACT-R

Cognitive architectures, such as ACT-R (described in Section 2.1, provide a framework for modelers to test computational models of cognition and to simulate human behavior. In this chapter, we use a constraint model and a drift diffusion model to inform the design of an extension to the ACT-R audio module for modeling attentional bias for spatial sounds. ACT-R has built-in modules for visual and auditory attention, which it communicates with through buffers [12]. For example, in the Visual module, a visual buffer contains all of the objects in the visual scene, and a visual-location buffer contains the location (as x and y coordinates) of the object that is currently being attended to. Although ACT-R has been used to test theories and simulate

tasks involving visual attention, auditory attention has received less consideration. ACT-R provides a basic audio module that assumes a constant amount of time to respond to sounds and provides no representation for spatial sounds. We extend this functionality to provide support for spatial sounds and the range of reaction times observed in a behavioral task.

3.3 Behavioral Experiments

We start by describing two behavioral tasks we designed to map out the attentional gradients.

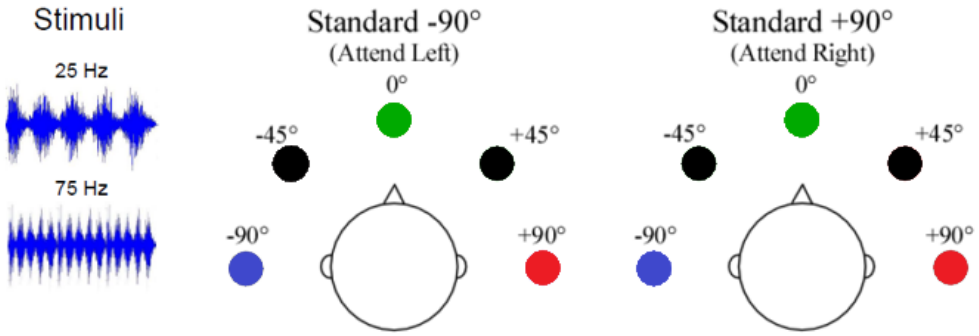

Figure 3.2: Experimental design for left (-90°) and right (+90°) conditions. Stimuli were presented from a standard attended location (-90°, 0°or 90°) or a shift location that was not currently being attended to (represented by the black circles). Subjects differentiated between stimuli with differing amplitudes (25 Hz vs 75 Hz).

3.3.1 Sustained Attention Task

In the first task, white noise was presented from five locations in the frontal plane ($-90°$, $-45°$, $0°$, $+45°$, $+90°$), and subjects (N=92) respond in each trial by discriminating the amplitude modulation (AM) rate, (25 Hz or 75 Hz). The slow AM rate sounds like a deck of cards being shuffled while the faster rate is perceived as a buzz. Subjects completed 6-minutes blocks with 150 trials coming from all five locations. In each block they attended to a standard location that was -90°, 0°, or

+90°(counterbalanced). Most stimuli came from a standard location, with probability 0.84, but sometimes shift to one of the remaining four distractor locations, with a probability of 0.04 per location.

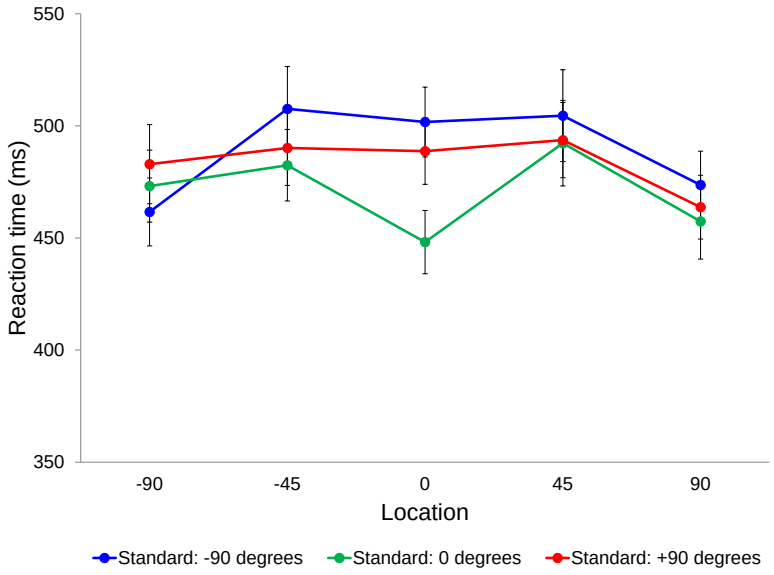

Figure 3.3: Median and standard deviation of reaction times in the spatial auditory attention task.

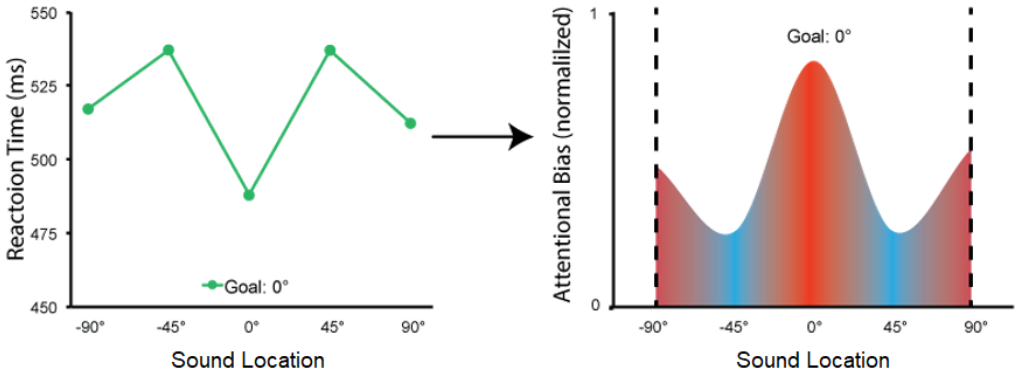

Figure 3.4: The relationship between reaction time and attentional bias. Attentional bias is on a normalized scale between 0 and 1. Faster reaction times mean higher attentional bias (shown in red), while slower reaction times mean bias values closer to 0 (shown in blue).

Figure 3.3 plots median reaction times and location for each standard condition. First, all conditions had slower responses to distractors vs. standards ($p < .001$),

indicating attention shift costs. This effect was more prominent the left ($-90°$) vs. the right ($90°$) standard ($p < .01$), suggesting that it is faster to shift auditory attention from right-to-left than from left-to-right. The $0°$standard has an increase at near $\pm 45°$ locations, similar to the left standard, but a decrease for the $\pm 90°$ locations, similar to the right standard ($p < .001$). In each condition, reaction times sped-up for the distractor locations furthest from the standard ($p < .001$) (for example, $-90°$ is the furthest distractor location from $90°$. This was seen in each subject's first block, so is not due to carry-over effects from previous standard locations. The faster responses at far distractors cannot be accounted for by a graded reduction in bias from the attended location. Instead, we hypothesize that bias contributed by a saliency map leads to the heightened bias to far distractors (see Section 3.4.1). Figure 3.4 shows how we theorize attentional bias to relate to reaction time.

3.3.2 Vigilance Task

Subjects are generally very accurate in performing the spatial attention task, and four of the five locations that are tested have relatively few trials (n=12 trials for each shift location). This is problematic for modeling with the drift diffusion model (described in Section 3.4.2), which requires examples of both correct and incorrect trials for the analysis. To address this, we used data from an experiment that was similar to the spatial auditory attention task [61]. In this task, subjects were given 1,908 trials in one block lasting 38 min 10 seconds. Of these trials, 1,800 trials were the same as in the sustained attention task when the standard location was at 0 (mid-line). There were 1,512 stimuli presented from the standard location, and for the infrequent shifts ($p = .04$/location) 72 trials were given at each of the four locations flanking the mid-line (45, 90). This vigilance data set yielded six-times more trials than in the sustained attention task described in section 3.3.1. Figure 3.5 plots median reaction times and location for the vigilance task each standard condition.

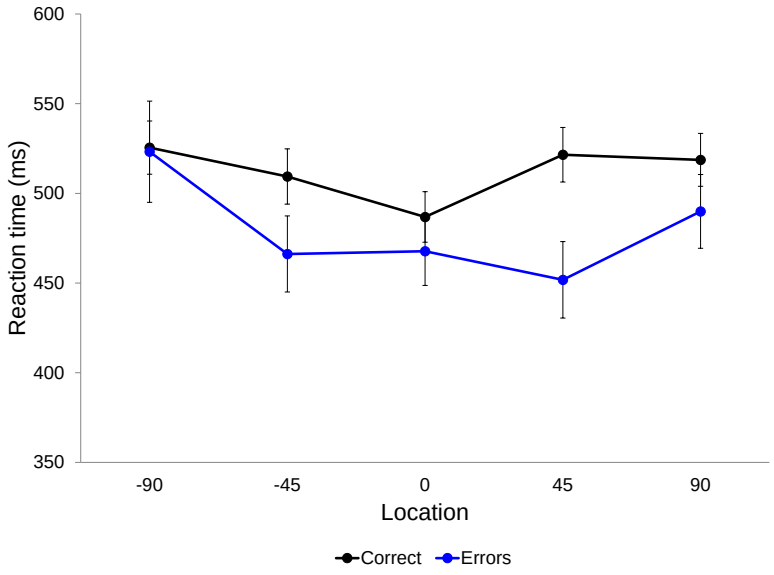

Figure 3.5: Median and standard deviation of reaction times for correct and error responses in the vigilance task.

3.4 Models of Spatial Auditory Attention

We examine two computational approaches to model spatial auditory attention. The first uses the constraint satisfaction framework to model how spatial auditory attention emerges from top-down and bottom-up interactions. The second uses the drift diffusion model to examine information accumulation in the behavioral task, showing how this leads to a sound being accurately perceived, or an error.

3.4.1 Constraint Model

Here we describe a constraint-based approach to modeling the cognitive mechanisms leading to attentional bias. We first describe the high-level details of how our model relates to spatial auditory attention and then present a formal description of the constraint model.

Figure 3.6 depicts the overall hypothesis on the interplay between top-down and

bottom-up spatial attention processing. There are three main components: a goal map that represents top-down attention, a saliency map that represents bottom-up attention, and a priority map that represents the combination of top-down and bottom-up attention. The given inputs to the model are (1) attended location and the (2) sound location. The output is a priority map representation of attentional bias across the 180° semicircle horizontal frontal plane (from −90° on the far left to 90° on the far right). Areas of greater attentional bias are assumed to relate to measurable data by having faster reaction times, more sensitive sensory thresholds, and increased accuracy relative to locations with less bias.

We emphasize that this is a model of information processing at the cognitive level. It is designed to help interpret behavioral results and inspire new experiments to test and refine the model. It is not intended to model how neural activity relates to attention. The gray boxes show inputs and outputs that interface with other cognitive functions.

We adopt a constraint-based representation that is very flexible in terms of modeling different hypotheses on the attentional bias distributions and on the interaction of the maps. Figure 3.7 depicts a high-level representation of the constraint graph of our model. Using this method, we were able to rapidly test combinations consisting of different goal map shapes and saliency map shapes. The goal map could be a Gaussian, consistent with a classic "attentional spotlight" or "zoom" lens gradient [62, 63], or a Gaussian flanked at the edges by inhibition commonly seen in visual attention studies [64], and supported by modeling and neurophysiological measures [65, 66, 67]. The saliency map could be either a constant bias at each location or a Gaussian spatially tuned to be opposite the goal map, with the highest attentional bias at the edges of the attended range.

We cast the interactions among the three maps into a constraint solving problem that can be efficiently solved with the rich algorithmic machinery which has been

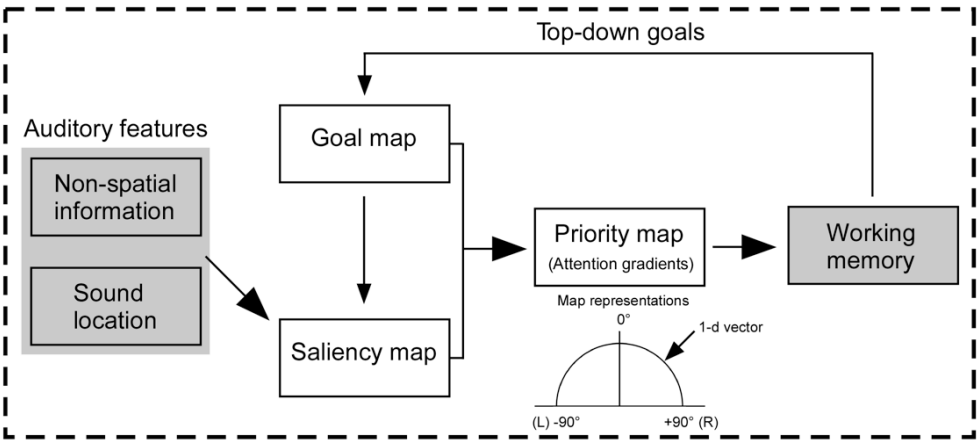

Figure 3.6: Constraint-based Computational Model Schematic.

developed for constraints[16]. A constraint satisfaction problem is defined as a triple $\langle X, D, C\rangle$ where X is a set of variables, $\{x_1, ..., x_n\}$; D is a set of domains, $\{D_1, ..., D_n\}$ associated with $x_1, ..., x_n$ respectively; and C is a set of constraints. In the notation below, a constraint $c \in C$ is a pair $c = \langle \sigma, \rho\rangle$ where σ is a list of variables and ρ is a list of functions defining the simultaneous variable assignments that are allowed by the constraint for the variable in σ.

In our constraint satisfaction problem we define variables in X, and domains in D are defined each with their own domain in D:

- L represents the attended location, with the domain being locations (in 1° increments) in the semicircle $\{-90°, -89°, ..., 89°, 90°\}$

- $V_G = \{V_G^i, ..., V_G^n\}$, where $i = \{-90, -89, ..., 89, 90\}$. V_G represents the *goal map* in the horizontal frontal plane (from -90°on the left to 90°on the right). The domain of the variable $V_G^i \in V_G$ is the interval $[0, 1]$ to represent attentional bias contributed by the *goal map*.

- $V_S = \{V_S^i, ..., V_S^n\}$, where $i = \{-90, -89, ..., 89, 90\}$. V_S represents the *saliency map* in the horizontal frontal plane. The domain of each variable $V_S^i \in V_S$ is the interval $[0, 1]$ to represent attentional bias contributed by the *saliency map*.

- $V_P = \{V_P^i, ..., V_P^n\}$, where $i = \{-90, -89, ..., 89, 90\}$ represents the *priority map* in the horizontal frontal plane. The domain of the variable $V_P^i \in V_P$ is also $[0, 1]$ to represent the total attentional bias in the *priority map*.

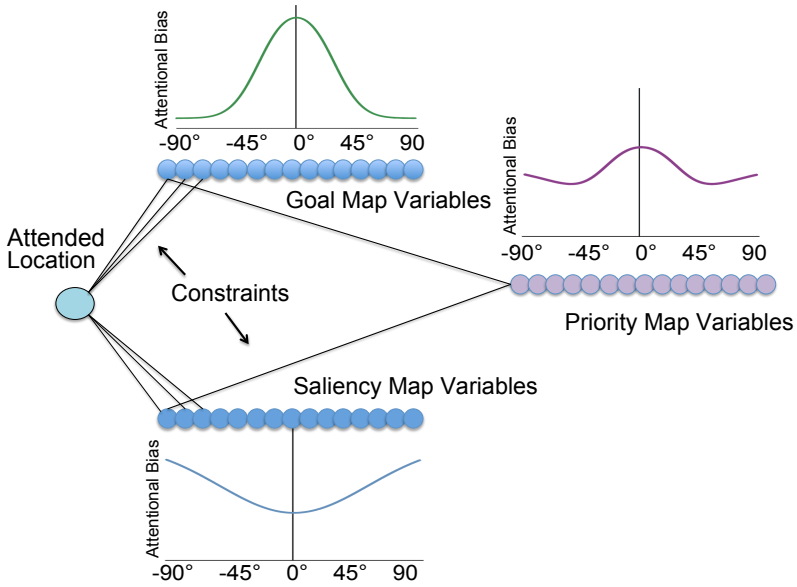

Figure 3.7: Variables and constraints representing the three maps and their interconnections. For clarity, only the constraints relative to the variables corresponding to the [-90°, -89°] location are shown.

Now we formally define the constraints that represent several different hypotheses about the shape of the goal map and saliency map and how they combine to produce attentional bias in the priority map. We denote variables in the goal map as V_G^i, the variables in the saliency map as V_S^i, and variables the priority map as V_P^i, where the i represents a 1° location in the azimuth plane.

Goal Map: Gaussian Model. This goal map represents a top-down, voluntary focus of attention to a location that has a progressive, symmetrical decrease in attentional bias away from an attended location. Given that location $L = l$ is (voluntarily) attended, this is represented as a standard Gaussian distribution using the following set of constraints over variables L and V_G^i. The following indicates the tuple of values

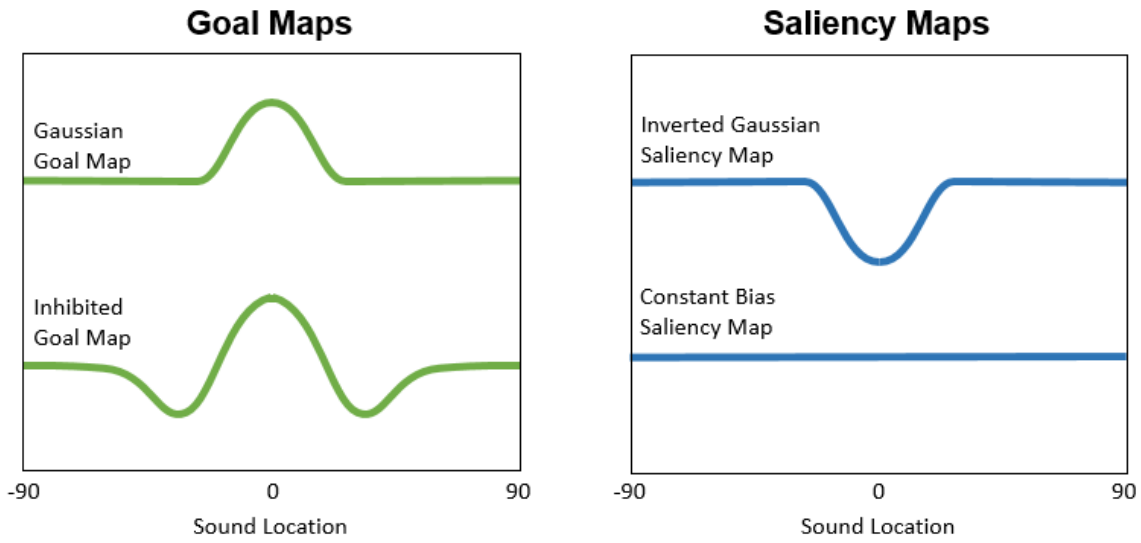

Figure 3.8: Hypothesized shapes for the goal map and saliency map.

which are allowed by the constraint.

$$\langle (L, V_G^i), (L = l, V_G^i = G_G e^{\frac{-(l-i)^2}{2d_G^2}}) \rangle \tag{3.1}$$

where d_G is the standard deviation of the goal map and G_G is the height of its peak (pictured in Figure 3.8).

Goal Map: Gaussian Model with Inhibition. This goal map represents a large amount of attentional bias at the goal location. But, rather than a systematic decrease in bias, this model examines the idea that attention is inhibited at the edge of the spatial range, leading to a "Mexican hat" shape (depicted as the *Inhibited Goal Map* in the bottom-left of Figure 3.8).

$$\langle (L, V_G^i), (L = l, V_G^i = G_G e^{\frac{-|l-i|^2}{2d_{G_1}^2}} + (G_G - G_G e^{\frac{-|l-i|^2}{2d_{G_2}^2}})) \rangle \tag{3.2}$$

Notice that this is obtained through the sum of a Gaussian and an inverted Gaussian. G_G is the maximum of the two functions, and there are two standard deviations for the components, represented by d_{G_1} and d_{G_2}. Using this approach, we obtain the

desired shape, which has a peak at the attended location, which dips down to an area of lower attentional bias and then increases and stabilizes as we move far away from the attended location.

Similarly, we consider two models of the saliency map, which models bottom-up attention, allocating attention to a stimulus based on how salient it is. As in the goal map, this involves defining constraints between variable L and the variables representing attentional bias at each location in the saliency map.

Saliency Map: Inverted Gaussian Model. In this model, bottom-up attention is reciprocally tuned away from the goal map, with less bottom-up attention being allocated at the attended location and more at locations further from it (pictured as *Inverted Gaussian Saliency Map* in the top-right of Figure 3.8). This can be represented as an inverted Gaussian distribution, represented by:

$$\langle (L, V_S^i), (L = l, V_S^i = G_S - G_S e^{\frac{-|l-i|^2}{2d_S^2}}) \rangle \tag{3.3}$$

where d_S is the standard deviation for the saliency map, and G_S is its maximum value.

Saliency Map: Constant Bias Model. We also tested the possibility that the saliency map is a constant level k of bias across all spatial locations (pictured in Figure 3.8). This is represented as:

$$\langle (L, V_S^i), (L = l, V_S^i = k) \rangle \tag{3.4}$$

where k is a constant value.

Priority Map. Lastly, the priority map is defined as the sum of the contributions of the goal and saliency map.

$$\langle (V_G^i, V_S^i, V_P^i), (V_G^i = u, V_S^i = v, V_P^i = u + v) \rangle \tag{3.5}$$

Using the models for the goal and saliency maps described above, we tested combinations that represent three different hypotheses about attentional bias in spatial auditory attention. These included: (1) Simple Gaussian Model: Gaussian goal map and constant saliency map, (2) Inhibited Goal Map Model: Gaussian goal map with edge inhibition and constant saliency map, and (3) Reciprocal Model: Gaussian goal map and a Gaussian saliency map tuned to opposite the goal map. Both (2) and (3) test whether attention gradients need to be more complex than a simple Gaussian, to account for our data.

3.4.2 Drift Diffusion Model

We applied the drift diffusion model to examine how information accumulates over time in the spatial auditory attention task. The drift diffusion model is commonly used to model the subject's information accumulation processes in two-choice response tasks, explaining performance differences by considering the speed and accuracy trade-off [17].

The model includes several components. First, the evidence is accumulated towards either choice in a noisy, stochastic way, with the average rate of the accumulation set to the drift rate (λ). Each choice is specified at a boundary (a), which specifies the threshold of information that must be accumulated towards one choice or another to make a decision. A parameter, z, represents the starting point where information begins to accumulate. The non-decision time (T_{er}) represents early perceptual encoding before the decision processes and the time to complete the action once a decision has been made.

There are a variety of computational methods for fitting the drift diffusion model to behavioral data [9]. These methods generally require estimating the parameters that best fit the reaction time distribution of both the accurate responses and errors. This requires a full reaction time distribution as input (including error responses).

However, in many behavioral tasks (including the ones described in Sections 3.3.1 and 3.3.2), the data contains few errors, particularly for responses at each of the infrequent shift locations. A small error reaction time distribution can result in inaccurate parameter estimations by the drift diffusion model.

To address this limitation, we used the EZ-Diffusion model to estimate the parameter values. The EZ-Diffusion model, which is derived from the Ratcliff drift diffusion model, is an alternative for sparse datasets [68]. Rather than requiring the full reaction time distribution, the EZ-Diffusion model only requires the mean (m) and variance (v) of response times and response accuracy (P_c) as inputs. The EZ-Diffusion model assumes that $z = a/2$ and translates the drift rate (λ), the boundary (a), and the non-decision time (T_{er}) from the inputs, using the following equations.

First, the probability that the stochastic process reaches the correct boundary, leading to a correct response, P_c, is:

$$P_c = \frac{1}{1 + e^{-av}} \tag{3.6}$$

Second, the variance ($v_r t$) of a symmetrical diffusion process [69] is calculated as:

$$v_{rt} = (\frac{a}{2v^3})\frac{2ye^y - e^2y + 1}{(e^y + 1)^2}, \tag{3.7}$$

where $y = -va$ and $v \neq 0$. If $v = 0$, then:

$$v_{rt} = a^4 \tag{3.8}$$

Finally, the mean reaction time is made up of two components, the mean decision time (m_{dt}) and the non-decision time (T_{er}) [68]:

$$m_{rt} = m_{dt} + T_{er} \tag{3.9}$$

Here, the mean decision time can be determined by the equation:

$$m_{dt} = (\frac{a}{2v})\frac{1-e^{y}}{1+e^{y}}. \tag{3.10}$$

Given the above equations, we can determine the values of a, T_{er}, and λ:

$$\lambda = \frac{|P_c - \frac{1}{2}|}{P_c - \frac{1}{2}} \left(\frac{\log \frac{P_c}{1-P_c}({P_c}^2 \log \frac{P_c}{1-P_c} - P_c \log \frac{P_c}{1-P_c} + P_c - \frac{1}{2}}{v_{rt}} \right)^{\frac{1}{4}} \tag{3.11}$$

$$a = \frac{\log \frac{P_c}{(1-P_c)}}{\lambda} \tag{3.12}$$

$$T_{er} = m_{rt} - m_{dt} \tag{3.13}$$

Using the the mean reaction time, variance and accuracy of the each subject's responses to the task described in Section 3.3.2, we used the above equations to calculate the parameter values a, λ and T_{er} (see Section 3.5.2).

3.5 Methods

In the following, we describe the methods utilized for analyzing the constraint model and drift diffusion model.

3.5.1 Constraint Model

By comparing the output of the constraint model to that of behavior in the spatial auditory attention task (described in Section 3.3.1), we can see how it explains the interplay of top-down and bottom-up behavior. To validate this model, we employed non-linear least-squares analysis (as implemented in the Python *scipy.optimize* library) to find the parameter values for d_S, d_G, G_S, and G_G that led to an optimized

fit when comparing the priority map to the data. Bootstrapping methods were used to compare model fit to 100 subsets of the data (without replacement), for each standard location. This was used to assess the consistency of results and prevent overfitting. For each bootstrapping run, half of the subjects (n=46) were randomly selected to train the model. Once the model was trained, we calculated the fit of the parameters using the root-mean-square error on the mean of the remaining subjects (n=46).

The parameter values representing the best fitting models are described in Section 3.6.1. These represent the best shapes for a goal and saliency map (representing top-down and bottom-up attention) that combine to create observed behavior in spatial auditory attention. Next, we consider how attention affects the information accumulation process and how this leads to a sound being perceived, or an error.

3.5.2 Drift Diffusion Model

Analyzing the parameters extrapolated from the collected data in the vigilance task allowed us to better understand how subjects accumulated information throughout the task. First, we estimated the drift diffusion model from the group of 30 subjects that performed the vigilance task described in Section 3.3.2. Subjects continuously attended to the midline location for over 38 min, and the inter-stimulus interval was faster (1.2 seconds between stimuli), which yielded six-times more trials than in the sustained and divided attention conditions above. The reason for needing more data is that estimating parameters for the diffusion model requires a dataset that contains correct and incorrect trials for each level of analysis. Subjects are generally very accurate in performing the spatial attention task, and four of the five locations that are tested have relatively few trials in the first two groups of subjects (n=12 trials for each shift location). Here subjects received 1,800 trials, with 72 trials for each of the four shift locations. One participant did not have enough errors to compute the diffusion parameters. Using the behavioral data for the remaining 29 subjects, we used

the EZ-Diffusion Model to estimate the values of the drift rate, decision boundary, and non-decision time. An analysis of the best fitting parameters is described in 3.6.2.

3.6 Results and Discussion

This following subsections discuss the results of analyzing spatial auditory attention using the constraint model and drift diffusion model.

3.6.1 Constraint Model

We fit the three combinations of the goal maps and saliency maps to the reaction times from our behavioral task when attending to the left, midline, and right locations, using the approach described in Section 3.5.1. The results are shown in Figures 3.9, 3.10, and 3.11. Parameter and fit values for each model are summarized in Tables 3.1, 3.2, and 3.3. These models ranged in complexity from simple to more complex. The simplest model (Simple Gaussian Model) featured a Gaussian goal map (see Equation 3.1) and a constant value for the saliency map (see Equation 3.4). In the second model (Inhibited Goal Map Model), the saliency map is also represented by a constant value, and the goal map is represented as a Gaussian peak at the location of top-down attention focus is surrounded by areas of attentional inhibition at the distant locations (see Equation 3.2). Finally, the most complex model (**Reciprocal Model**) featured opposing Gaussian functions for goal map and saliency map. Here, the goal map peak is at the top-down attended location. The saliency map peak is centered away from the goal map at the most distant location from the goal map focus (see Equation 3.3). For example, when the goal map is centered on -90, the attentional bias contributed by the saliency map peaks at +90.

The results show that among the three models being tested, a simple Gaussian attention gradient that decreases with distance from the attended standard location

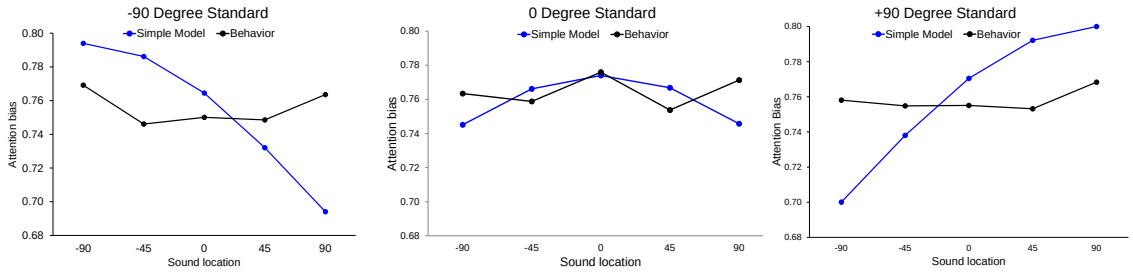

Figure 3.9: Simple Gaussian Model: Gaussian Goal Map and Constant Saliency Map. Comparison of the predicted and actual attentional bias as a function of attended standard and stimulus locations.

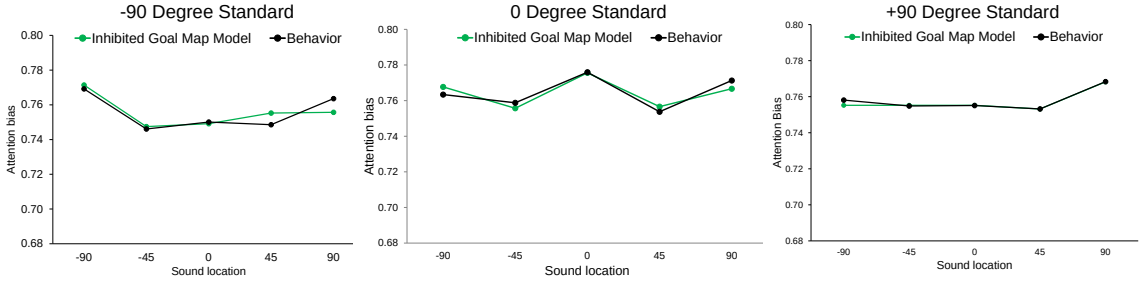

Figure 3.10: Inhibited Goal Map Model: Inhibited Goal Map and Constant Saliency Map. Comparison of the predicted and actual attentional bias as a function of attended standard and stimulus locations.

had the worst fit to the behavioral data in (Figure 3.9). Statistical comparisons of the fit showed significantly worse fits relative to the Reciprocal Model (all 3 standards, $p < .001$) and the Inhibited Goal Map Model (all 3 standards, $p < .001$). The results of quantitative tests in the present task show that one of the most common metaphors for the shape of spatial attention; that of a spotlight or zoom lens does not account well for the present data.

We next considered the Inhibited Goal Map Model, which featured a more complex goal map shape and a zone of inhibition at the edges of the gradient. This shape is supported in some visual attention studies [64, 65]. In this approach, the saliency map is still considered as a constant level of bias in all directions within the range of tested sound locations. This model fits our behavioral data well, as seen in Figure 3.10.

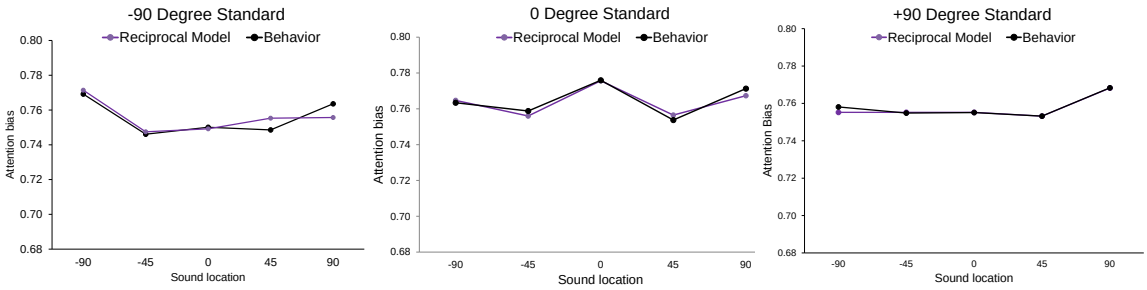

Figure 3.11: Reciprocal Model: Gaussian Goal Map and Inverted Gaussian Saliency Map. Comparison of the predicted and actual actual attentional bias as a function of attended standard and stimulus locations.

The Reciprocal Model (pictured in Figure 3.11) was also a good fit for the behavioral data. Here the goal and saliency maps are both modeled using Gaussian functions, with the saliency map being inverted so that it is tuned away from the focus of attention for the goal map. Pairwise t-tests comparing the fit over 100 runs, showed that the mean fits did not differ among any of the standard locations ($p > .10$). Taken together, the above results show that a typical linear or Gaussian-shaped attention gradient is insufficient to explain our reaction time results. The two models that were viable either had a more complex gradient shape for the goal map (by combining the Inhibited Goal Map with the Constant Saliency Map) or reciprocal Gaussian shapes for both the goal and saliency maps.

		Level		SD
Standard Location	Fit	GM	SM	GM
-90°(left)	0.0039	0.30	0.49	200.0
0°(midline)	0.0018	0.30	0.48	200.0
90°(right)	0.0051	0.30	0.50	200.0

Table 3.1: Simple Gaussian Model. Best fitting Goal Map (GM) and Saliency Map (SM) parameters and resulting fits at each standard location. This model was comprised of a Gaussian shaped goal map and constant value for the saliency map.

		Level		SD
Standard Location	Fit	GM	SM	GM
-90°(left)	0.0016	0.40, 0.42	0.36	59.2, 68.8
0°(midline)	0.0006	0.38 ,0.45	0.40	50.0, 60.1
90°(right)	0.0006	0.45, 0.44	0.32	51.6, 53.2

Table 3.2: Inhibited Goal Map Model. Best fitting Goal Map (GM) and Saliency Map (SM) parameters and the resulting fits at each standard location. The model was comprised of a goal map with Gaussian peak at the focus of top-down attention that was flanked by areas of attentional inhibition at the distant locations. The saliency map was represented as a constant value.

		Level		SD	
Standard Location	Fit	GM	SM	GM	SM
-90°(left)	0.0026	0.74	0.76	62.7	67.8
0°(midline)	0.0010	0.75	0.80	47.9	52.4
90°(right)	0.0008	0.75	0.74	42.0	45.0

Table 3.3: Reciprocal Model. Best fitting Goal Map (GM) and Saliency Map (SM) parameters and resulting fits at each standard location. In this model, the goal map peak is at the top-down attended location. The saliency map peak is centered away from the goal map at the most distant location from the goal map focus.

3.6.2 Drift Diffusion Model

Figure 3.12 plots the best fitting λ (drift rate), a (boundaries) and T_{er} (non-decision time) as a function of stimulus location. Separate one-way ANOVA tests found significant effects of location for drift rate ($p < .001$) and decision boundary ($p < .01$). Comparisons of shift locations using 2 (side) x 2 (eccentricity) ANOVAs found a significant effect of eccentricity for non-decision time ($p < .02$), with longer non-decision times at the most lateral shift locations ($\pm 90° > \pm 45°$). There was a similar trend towards faster drift rates for the most lateral shift locations ($\pm 90° > \pm 45°$, $p = .053$). Thus, relative to the shift locations, information accumulated faster at the standard location, decision thresholds for responding were higher, and non-decision processes took less time. Parameters at shift locations in the left and right hemispaces were comparable, and there were small differences in non-decision time, and perhaps drift rate, between $\pm 90°$ and $\pm 45°$ locations.

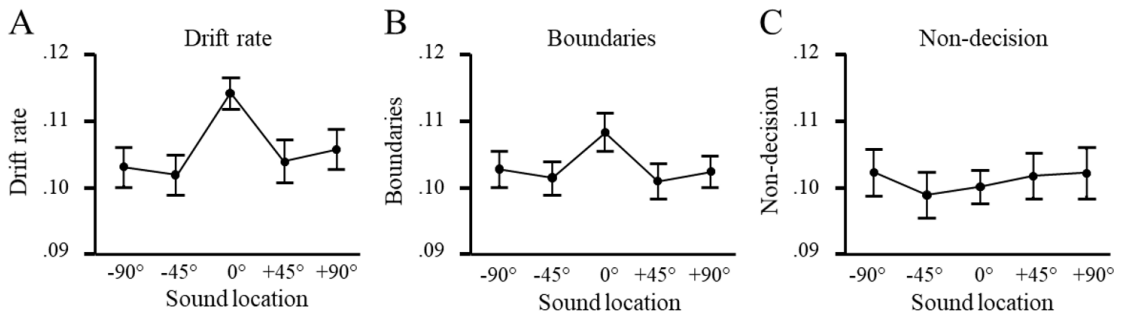

Figure 3.12: Parameters plotted against sound location. (A) drift rate λ, (B) boundary a and (C) non-decision time T_{er}

From these results, we can observe how the interplay between drift rates, decision boundaries, and non-decision time can relate to the overall attentional bias and accuracy. We see higher attentional bias and faster reaction times at the standard location, as well as the lateral shift locations ($\pm$90). Since people were more accurate at the standard location, we see more conservative decision thresholds, higher drift rates and less non-decision time. In the $\pm$45 and $\pm$90 locations, the accuracy was lower, leading to lower decision threshold boundaries. However, in the most lateral shifts at ($\pm$90), the bias and reaction times were faster, resulting in slightly increased drift rates and boundaries in the diffusion model. These results provide more insight into how the rate of information accumulation (drift rate), which is thought to be enhanced by selective attention [67], relates to the placement of the decision boundaries. The decision boundaries are the main mechanism for modeling speed-accuracy tradeoffs in diffusion models; liberal thresholds confer faster speed and lower accuracy, while conservative boundaries slow responding but improve accuracy [17].

3.7 Extending the ACT-R Audio Module

ACT-R includes a basic audio module that allows cognitive agents to attend and respond to simulated sounds. However, the existing module does not provide sup-

port for spatial sounds or the varied reaction times we observed at different spatial locations. This simplified view of auditory attention was not expressive enough to represent the effect of attentional bias or the range of reaction times observed in our behavioral data. In the following sections, I show how we used the constraint model and drift diffusion model to expand this module to support spatial sounds.

To create an environment for testing our model of spatial auditory attention, we extended the ACT-R audio module to model the attentional bias as a combination of top-down and bottom-up processes. Top-down processes are represented as target features (attended location l) stored in the ACT-Rs Imaginal buffer (see Section 2.1 for more information). Bottom-up processes are represented using features (sound location i) of sounds stored in ACT-R's Aural-Location buffer. Using these two buffers, we generate a priority map that is generated using the Gaussian Goal Map (Equation 3.1) and Inverted Gaussian Saliency Map (Equation 3.3), as described in Section 3.4.1.

After the map of attentional bias is calculated, it must be converted into the amount of time spent attending to a sound. A timeline of the components making up the total response time is illustrated in Figure 3.13. ACT-R provides some built-in timings for detecting a sound (50 ms), encoding a tone (50 ms), identifying an appropriate production rule that determines which key to press (50 ms), and pressing the key (160 ms). The amount of attentional bias for a sound at location i is represented by V_{P^i}. This value represents the inverse of the total reaction time, represented by the equation $2000(1 - V_{P^i})$. This provides a value representing the number of milliseconds between 0 and 2000. Since 310 ms is already accounted for by other ACT-R modules, we take the resulting reaction time and subtract 310. The resulting value is used as the time in milliseconds that the model will spend attending to a detected sound before choosing which key to press. Some sets of model parameters can result in a priority map that generates an attentional bias of 0.845 or greater, which will

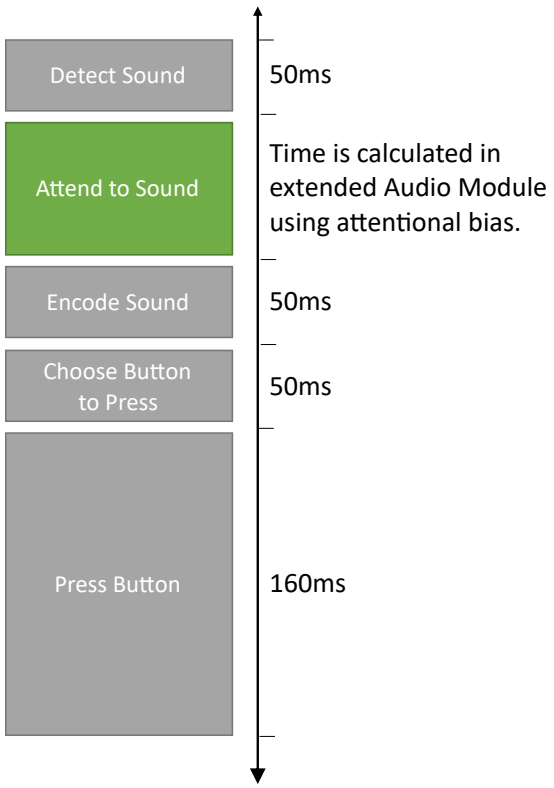

Figure 3.13: Amount of time (in ms) that each step contributes to the overall response time in ACT-R's simulated spatial auditory attention task.

result in negative reaction times. In such a scenario, the model will substitute a value of 0 ms. Thus, the possible reaction time required for the ACT-R model to complete this task will be between 310 ms and 2000 ms. The equation for calculating the time contributed by attending to a sound, given the attentional bias b at sound location i, is represented below.

$$t(V_{P^i}) = \begin{cases} 2000(1 - V_{P^i}) - 310 & V_{P^i} < 0.845 \\ 0 & V_{P^i} >= 0.845 \end{cases}$$

3.7.1 Constraint Model Implementation in ACT-R

With this extended audio module, modelers can create an ACT-R agent that mimics human subjects completing the spatial auditory attention task. To simulate the behavioral task, the cognitive agent is instructed to attend to either the $-90°$, $0°$ and $90°$ standard locations and then respond to sounds presented from the five sound locations that the human subjects also responded to (at $-90°$,$-45°$,$0°$,$45°$, and $90°$). The agent chooses a response action depending on the AM-rate of the sound (25 Hz or 75 Hz). This involves invoking the Motor module to simulate pressing the appropriate key. The response time varies based on the amount of attentional bias at the sound location.

An attended location can be modeled in ACT-R by storing the location (in degrees) in the Imaginal buffer, which is the buffer that ACT-R uses for storing task-relevant information. Using this information and the sound location (stored in the Aural-Location buffer), the agent chooses how to respond using the production rules. For this task, four production rules are sufficient to govern the agent's behavior. These include:

1. If a location is in the Imaginal buffer, try to find a sound in the environment and move it to the Aural-Location buffer.

2. If a sound is detected in the Aural-Location buffer and a location is in the Imaginal buffer, encode the sound and move it to the Aural buffer.

3. If the sound has an AM-rate of 25hz, then press "d".

4. If the sound has an AM-rate of 75hz, then press "k".

Given the attended location in the Imaginal buffer and the sound location from the Aural-Location buffer, the Audio module calculates the attentional bias and converts this to the appropriate response time in milliseconds as described above. The ACT-R simulated data compared to the behavioral data can be seen in Figure 3.14.

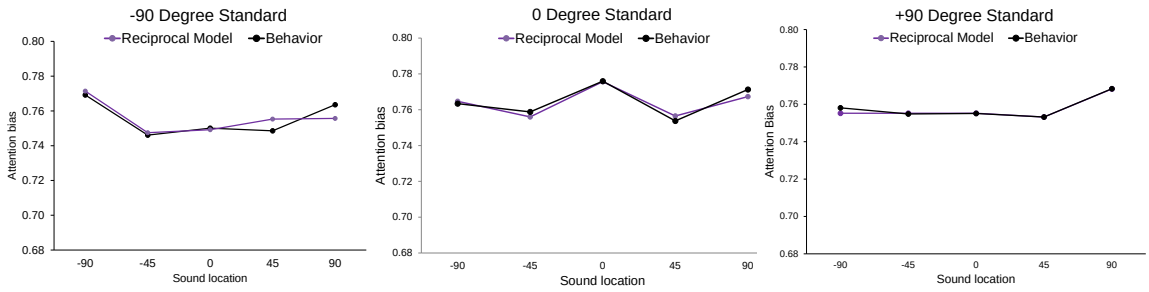

Figure 3.14: Comparison of the mean of 100 response times simulated in ACT-R and the mean response times of the 92 subjects in the sustained attention task. We recall that the Reciprocal Model, defined in Section 3.4.1, is used by ACT-R to generate these response times.

3.7.2 Modeling Individual Differences and Errors

Ideally, our ACT-R simulation should (1) generate a range of realistic reaction times, and (2) simulate errors that are in line with observed behavior. By default, the ACT-R audio module calculates an average reaction time and does not predict errors. We addressed this in two ways, including the introduction of noise parameters to generate reaction times in an exGaussian distribution and 2) by simulating response times and errors using the drift diffusion model.

ExGaussian Distribution. As is often the case in many psychological experiments that measure reaction time [70], the reaction time data collected in the experiment described in Section 3.3.1 is right-tailed, appropriate for modeling as an

exGaussian distribution. ExGaussian distributions are the convolution of a normal distribution (with a mean of μ and a standard deviation of σ) with an exponential distribution (with a mean of τ), expressed as.

$$f(t) = \frac{1}{\tau} exp\{-\frac{(t-\mu)}{\tau} + \frac{\sigma^2}{\tau^2}\} * \phi\{\frac{(t-\mu)}{\sigma} - \frac{\sigma}{\tau}\}, \tag{3.14}$$

where ϕ is a normal CDF.

We added parameters to the audio module that allow the modeler to set the value of σ and τ. μ is automatically set to the value generated by the constraint model. Using these three parameters, the module generates reaction times in an ex-Gaussian distribution, allowing the modeler to simulate the range of reaction times observed in an experimental task.

To generate a distribution of reaction times that were similar to those observed in the behavioral experiment, we employed non-linear least squares analysis (as implemented in the Python *scipy.optimize* library) to find the parameter values function provided in the Python scipy library to find the the parameter sets $[\sigma_s(a), \tau_s(a), \sigma_d(a), \tau_d(a)]$ that best fit an exGaussian distribution to each individual subject's reaction time at each attended location, a. The parameters $\sigma_s(a)$ and $\tau_s(a)$ represent the best fitting values for σ and τ at the standard location, a. $\sigma_d(a)$ and $\tau_d(a)$ represent the best fitting parameter values at unattended locations (every location except a). This resulted in a sample of 92 sets of parameters that represent the exGaussian distributions of every subject at each attended location, $[-90°, 0°, 90°]$.

These parameters were used to modulate the exGaussian noise for each subject simulated by ACT-R. To simulate the subjects in our behavioral task, we randomly chose a set of parameter values (p_i) from the 92 to use for each subject. The ACT-R audio module calculates the amount of time required to attend to a sound, using the method described in 3.7.1. The resulting time (in milliseconds) is used to fix the value of μ in an exGaussian distribution for the current attended location l and

the location of sound i. If the incoming sound is coming from the attended location ($l = i$), then the values of $\sigma_{standard}$ and $\tau_{standard}$ from p_i are used to fix the remaining σ and τ parameters in the exGaussian distribution. If the incoming sound comes from a location that is not attended, then $\sigma_{deviant}$ and $\tau_{deviant}$ are used. A random time is chosen from the resulting exGaussian distribution and used as the simulated response time. Using this method, it is possible to simulate a range of reaction times with a similar mean and standard deviation to that of the behavioral data.

Drift diffusion process In addition to simulating a range of reaction times, a realistic representation requires ACT-R agents to also simulate errors. The drift diffusion model provides a means to simulate both a range of reaction times and error rates in binary choice tasks.

We used the Wiener method, a common method for simulating diffusion processes [9], to implement the drift diffusion model in our Audio Module extension. Additionally, we added a module parameter that allows a modeler to choose to bypass the constraint model and use the drift diffusion model instead. When this parameter is set to True, the Audio Module will calculate reaction times and errors with the drift diffusion module.

When the drift diffusion model is enabled, the modeler must supply values that represent the drift rate λ, decision threshold boundary a and non-decision time, T_{er}. Given these values, the module will simulate the information accumulation process by randomly accumulating values towards 0 or boundary a until crossing the threshold, adding a small amount of time to the deliberation process at each step. If the information accumulation process reaches a, then a correct response will be processed, the frequency of the sound is encoded correctly, and the Motor module chooses the correct button to press. If the accumulated evidence trends towards 0 and eventually passes this threshold, then an error is recorded, the frequency of the sound will be incorrectly encoded, and the Motor module will simulate an incorrect button press.

Using the drift diffusion method, we completed 100 simulations of the vigilance task in ACT-R. For each simulation, we used a set of drift diffusion parameters (λ, a and T_{er}) randomly chosen from the parameters that were calculated using the EZ-diffusion method described in Section 3.4.2. This resulted in a reaction time and error distribution that is very similar to the behavior we observed in the sustained attention task and vigilance task ($R^2 = 0.9996$, $RMSE = 8.5$). A comparison of the simulated data to the behavioral data is shown in Figures 3.15 and 3.16.

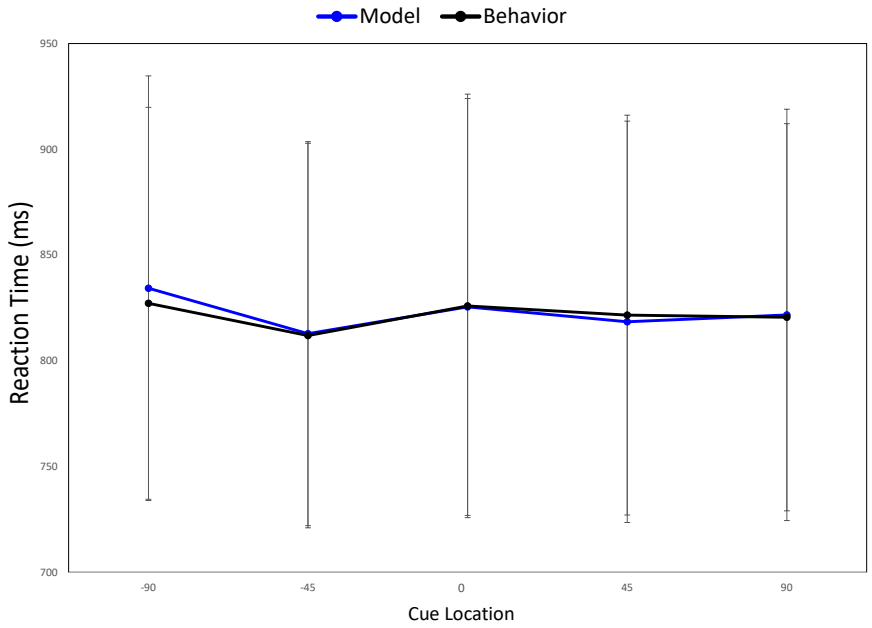

Figure 3.15: ACT-R simulation results (using the drift diffusion model) reaction times compared to behavioral data.

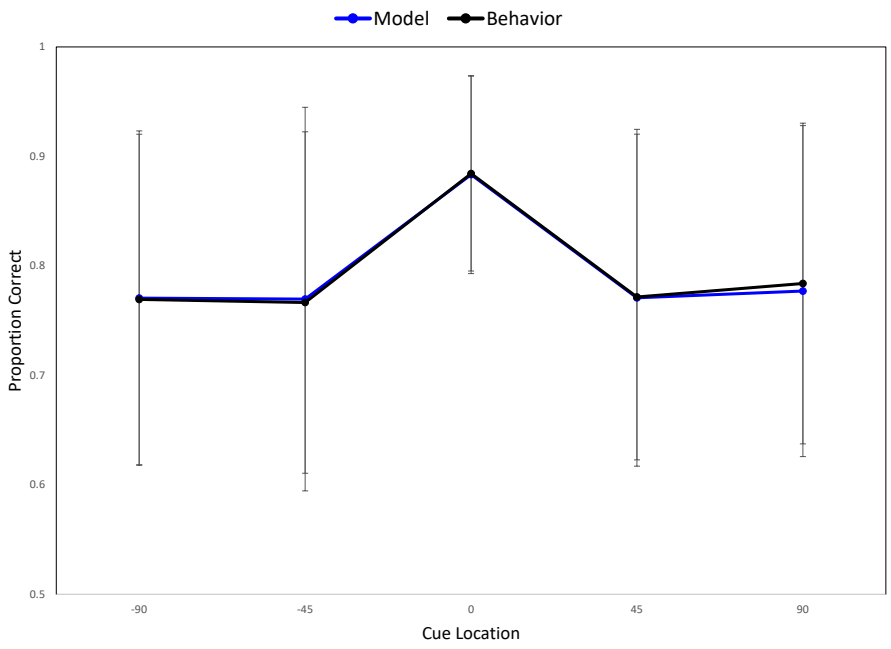

Figure 3.16: Model (using the diffusion method) accuracy compared to behavioral data.

3.8 Conclusion and Future Directions

In this chapter, we presented two new approaches to model attentional bias in spatial auditory attention, using the results of two behavioral tasks designed to elicit this type of bias. In the first approach, we used the framework of constraint satisfaction problems to rapidly test different combinations of three different goal maps and three different saliency maps. We compared how well each combination fit the behavioral data. We found two combinations to be effective in modeling the behavioral data, including the Reciprocal Model and Inhibited Goal Map Model, described in Section 3.6.1. These represent two possible hypotheses about the shape of the saliency map. First, the saliency map may contribute an equal amount of attention across the attention range, with top-down inhibition being the primary driver of the Mexican hat shape observed in the behavioral data. Second, the saliency map may be an inverted Gaussian shape, contributing more attention near the edges of the range.

Further experimental work must be done to determine the shape of the saliency map. It would also be interesting to examine how the saliency map and resulting priority map change as the range of attention increases from 180°to 360°.

We supplemented the results of our constraint model by using the EZ-diffusion model to model how information accumulates in this spatial auditory attention task. We showed how EZ-diffusion parameters, including the drift rate and boundary separation, were related to the attentional bias and accuracy at each spatial location. By using a diffusion model, we were also able to investigate how these parameters relate to whether a subject will correctly perceive a sound or respond in error.

Finally, we presented an extension to the ACT-R audio module that uses the constraint model to predict response times to cues that are presented from different spatial locations. The extension incorporates parameters that allow modelers to model individual differences in the spatial auditory attention task. One approach allows simulating response times as an exGaussian distribution, while the second model's individual differences in response times, as well as perceptual errors in binary choice tasks, using the drift diffusion model. Continued development on the audio module represents an interesting area of continued research that will open up new possibilities for improving our understanding of spatial auditory attention and modeling tasks where audition is important.