Modeling spatial auditory attention: handling equiprobable attended locations

Jaelle Scheuerman
Tulane University
New Orleans, LA, USA
jscheuer@tulane.edu

K. Brent Venable
Tulane University
New Orleans, LA, USA
kvenabl@tulane.edu

Maxwell T. Anderson
University of Texas
San Antonio, TX, USA
maxwell.anderson@utsa.edu

Edward J. Golob
University of Texas
San Antonio, TX, USA
edward.golob@utsa.edu

Abstract

Attention has been the focus of a considerable amount of research in cognitive models. Yet, most of the work has been devoted to studying visual attention. In this paper we focus, instead, on auditory attention and on a model for how it is distributed in space following basic ideas of top-down and bottom-up attentional control from verbal models. In particular, we extend a previous computational model [Golob *et al.*, 2016; 2017] which is organized around three main components: a goal map, a saliency map, and a priority map. The goal map models the distribution of attention which is allocated by choice (top-down component). The saliency map, as the name suggests, models attention related to the saliency of auditory stimuli (bottom-up component) and the priority map synthesizes the other two maps in an overall distribution of the attentional bias. This model was shown to be successful in modeling behavioral data of experiments where there is a single attended location. We relax this assumption and extend the framework to encompass scenarios where there can be multiple attended locations. Most importantly, we leverage the parameters learned by fitting the behavioral data with single attended location to make predictions for the case in which sounds are presented at multiple locations with equal probability. Our predictions feature a very small error with respect the new behavioral data and are shown to leave very small room for improvement. This is an important step in the, still largely unexplored, field of auditory attention modeling as it provides a first example of how the computational model can be used as a predictor.

1 Introduction and Motivation

The auditory system is differentiated from other senses in that it allows us to monitor the environment for sounds all around us, including those at a distance and out of sight. This allows us to sense and quickly shift attention to events, such as a predator snapping a twig behind us. Although spatial hearing promotes survival, it is a challenge to strike a balance between focusing on our current task and recognizing and shifting attention to threats or opportunities in the environment.

While visuospatial attention has attracted a substantial amount of attention from the research community [Cave and Bichot, 1999; Greenwood and Parasuraman, 1999; Itti and Koch, 2001], much less work has been devoted to model auditory spatial attention. Our purpose is to to better understand this at the cognitive and neural levels of analysis.

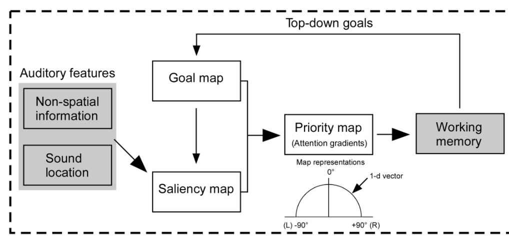


Figure 1: Computational Model Schematic.

Our approach consists in an interdisciplinary effort that uses behavioral methods to test and refine a computational model of spatial auditory attention and map out auditory attention as a gradient over space. Broadly speaking, our behavioral tasks measure reaction time to auditory stimuli generated at different locations in space. In previous work, we proposed a computational model of the interplay between top-down and bottom-up spatial attention. The model was tested using a task where sounds predominantly come from one location in space [Golob *et al.*, 2016; 2017]. In this paper we extend our model to handle scenarios in which attention is divided over multiple locations in space with similar probability. We base our extension on utilizing a combination of the models learned for focusing attention on a single location and show that this produces a good prediction when attention is divided over a range of locations.

1.1 Applications to Human-Centered Design

We foresee that this work will help advance understanding of basic issues in attention, such as top-down and bottom up interactions, vigilance and capacity limitations. It will help in identifying the implications of the auditory system's comparative advantage over other modalities in its ability to panoramically monitor the environment. Understanding these factors is imperative for designing systems for humans where audition is important. For example, it is an issue in aeronautic safety when pilots miss critical alarms in the auditory environment [Dehais *et al.*, 2014]. Similarly, clinicians find it challenging to learn and distinguish auditory alarms from medical devices [Edworthy *et al.*, 2017]. In both of these situations, it would be helpful to predict when such warnings might be ignored. Our work targets auditory spatial attention and thus it will play a fundamental role in understanding and enhancing the use of spatial sound in particular in virtual environments where it is already applied [Cohen *et al.*, 2015]. A computational model of spatial auditory attention will simultaneously bring insight on how the attention allocation processes work and play a significant role in facilitating and optimizing systems designed with audition in mind.

2 Background

Most psychological models of attention distinguish between attention that is guided by personal choice and that which is directed to a salient event, such as a loud sound [Pillsbury, 1908]. In literature, this can be referred to as top-down vs. bottom-up attention control. Top-down control biases attention towards information useful for fulfilling the current goals in short-term memory. On the other hand, bottom-up refers to how attention may be captured by something outside the top-down task set. This distinction is meaningful, although it is recognized that the two processes are highly interactive [Folk *et al.*, 1992] and can be challenging to distinguish between. This motivates our use of a computational model, using AI constraint programming, that allows examining top-down and bottom-up functions separately.

2.1 Auditory Attention

Attention can be expressed as a spatial gradient relative to an attended location [Cave and Bichot, 1999]. Gradients are likely reflect limitations in perceptual processing, but may also relate to limits in possible behaviors at a given moment [Allport, 1989]. Auditory spatial cuing decreases reaction times to subsequent targets at a cued location relative to uncued locations, as shown in [Zatorre *et al.*, 1999; Rorden and Driver, 2001], where target reaction times were found to increase monotonically with greater distance between the cued and target locations. Sometimes visual studies suggest that gradients may have a more complex shape, with reaction times increasing and then decreasing away from the cued location [Mller *et al.*, 2005; Caparos and Linnell, 2010](”Mexican-hat”). This is similar to our preliminary findings in the auditory modality, but the auditory results have a much larger spatial range.

2.2 Computational Models of Auditory Attention

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 [Itti and Koch, 2001; Lewandowsky and Farrell, 2010]. Our model uses basic ideas of top-down and bottom-up attention control from prominent verbal models [Baddeley, 2010; Cowan, 1988]. The novelty of the approach we consider here is the application to auditory spatial attention, which is not dealt with in detail in the general models. Our model is distinguished by focusing on auditory spatial attention and how it emerges from top-down and bottom-up interactions. Moreover, the models of attention mentioned above are designed as ad hoc mathematical descriptions of the considered phenomena, while we opt to cast our model into a more general artificial intelligence setting.

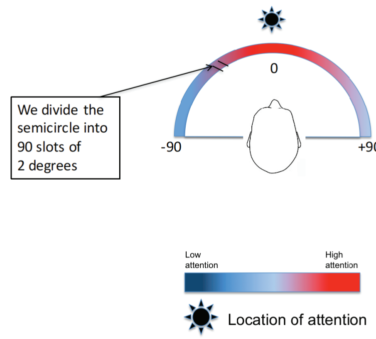


Figure 2: Map of attentional bias.

2.3 A Computational Model of Spatial Auditory Attention

In previous work, [Golob *et al.*, 2016; 2017] we presented an overall hypothesis of the interplay between top-down and bottom-up spatial attention processing. The model has three main components (white boxes in Figure 1): goal map, saliency map, and priority map. The gray boxes show inputs and outputs that interface with other cognitive functions.

Each map is a 1-D vector of attentional bias in normalized units (0-1) across the semicircular horizontal frontal plane (from -90°on the far left to +90°on the far right, in 2°increments, as shown in Figure 2). The goal map indexes top-down attention bias, and is a function of the central executive in verbal models. It models top-down, voluntary focus of attention to a location, and has a progressive, symmetrical decrease in attentional bias away from the attended location. The saliency map, instead, models how attention is allocated to a stimulus given how salient its characteristics are. The priority map synthesizes the contribution of the other maps. In all the maps 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.

This computational model adopts a constraint-based approach to cast the interactions among the three maps into a constraint solving problem. Constraint programming [Rossi *et al.*, 2006] is a powerful artificial intelligence paradigm for modeling and solving combinatorial search problems. The basic idea in constraint programming is that the user states the constraints and a general-purpose constraint solver is used to solve them. Constraints concern subsets of variables and define which simultaneous assignments to those variables are

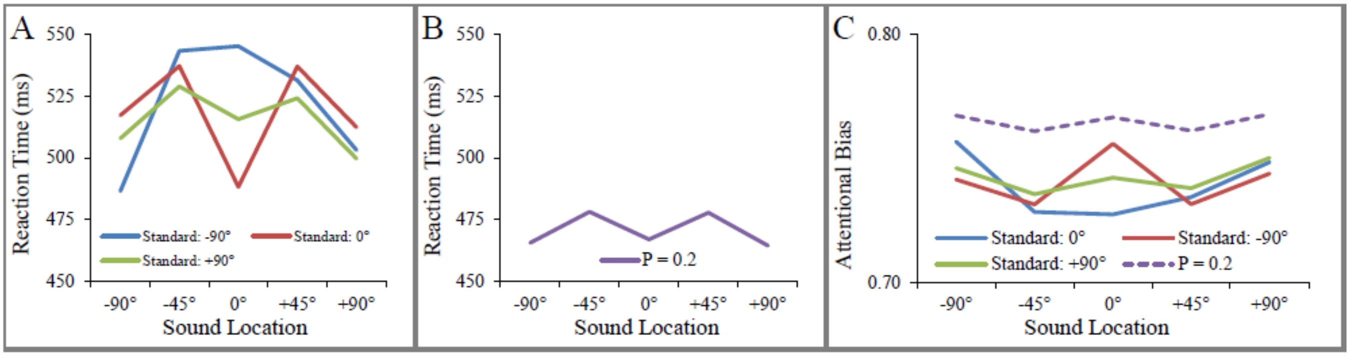


Figure 3: Behavioral data in the form of average reaction times at the five locations. Our first experiment showed distinctly different distributions for each of the three standards (A). In the new experiment, when sounds were equally likely from all locations, the distribution of average reaction times was significantly flattened (B). Part C displays the inverse of normalized mean reaction times for both experiments, showing how attention bias is theorized to relate to reaction time (more bias → faster reaction times).

allowed. In the model described in [Golob *et al.*, 2016], there is one input variable corresponding to attended location (A) with the domain being locations (2°increments) in the semicircle {-90,-88,...,0,...,88,90}.

Variables V_G^i, V_S^i and V_P^i represent respectively, the i-th variable of the goal, saliency and priority map where i ranges in {-90,...,+90}. The domain to quantify attentional bias uses normalized units (0-1, in .01 increments). The attentional bias in the goal map, given that location $A = a$ is (voluntarily) attended, is represented by a *standard Gaussian distribution* modeled as the set of constraints over variable A and V_G^i:

$$(A = a, V_G^i = G_G e^{\frac{-|a-i|^2}{2*d_G^2}}). \quad (1)$$

Here, d_G is the standard deviation of the goal map and G_G is the height of its peak.

Similarly the bias in the saliency map is constrained to a *inverted Gaussian distribution* by the following constraints:

$$(A = a, V_S^i = G_S - G_S e^{\frac{-|a-i|^2}{2*d_S^2}}). \quad (2)$$

Here, d_S is the standard deviation for the saliency map, and G_S is its minimum value.

Finally, the priority map is defined as the sum of the contributions of the goal and saliency map, with α and β between 0 and 1:

$$(V_G^i = u, V_S^i = v, V_P^i = \alpha u + \beta v). \quad (3)$$

This model was validated on results obtained from the behavioral task outlined in 2.4 by testing it against different options for the goal and saliency map. Equations 2 and 3 emerged as the best in terms of fitting the experimental data obtained from the behavioral task described in 2.4, for all three locations. The best fitting values $(1 - E(p))$, where $E(p) = \sum_{x \in \{-90°,-45°,0°,+45°,+90°\}} (d_x - p(x))^2$, were equal to: 0.943 for 0°, 0.739 for +90°and 0.904 for -90°.

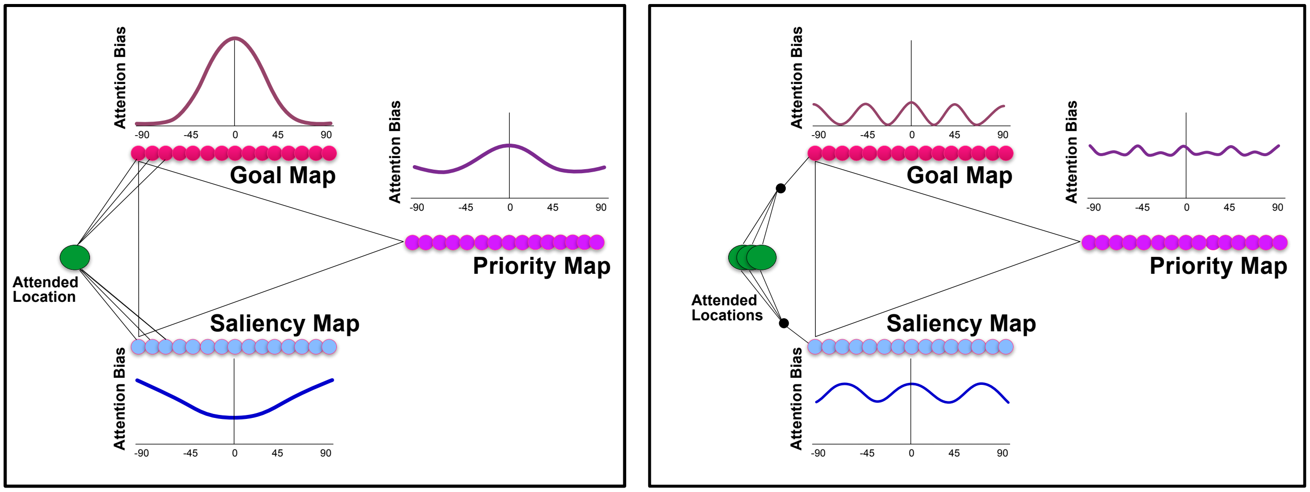
2.4 Behavioral task and results with a single attended location

We now describe the behavioral task generating the data we used to evaluate both our previous model and the new, more general, model presented in Section 3.

Study participants completed a simple behavioral task designed to index auditory attention across space. Participants were told to judge non-spatial aspects of sounds that came from different spatial locations. Individuals reaction times to sounds at 5 different locations in space were utilized to derive the relative distribution of attention over space. Two distinct types of white noise were presented from 5 possible locations across the participants' 180°frontal horizontal plane (90°, 45°, 0°, +45°, +90°). Subjects were instructed to discriminate between two types of noise via button press, with one button for each noise. Both noises were amplitude modulated at different rates (25Hz and 75Hz AM-rates). Noise with the 25Hz AM-rate was described to participants as a card shuffling sound, and noise with the 75Hz AM-rate was described to participants as a buzzing sound. Most stimuli came from a standard location (p = 0.84) but sometimes shift to a distractor location (p = 0.04). Separate blocks had the standard at -90°, 0°and +90°(counterbalanced).

Figure 3 (A) shows the average reaction time over 42 participants for the three standard locations. In Figure 3 (C), all lines represent the same results but inverted (using the formula $(2000 - x)/2000$ to show the attentional bias. Units of attentional bias are arbitrary, but correspond to the range of reaction times between 0 and 2000 ms. All conditions showed faster reaction times at the attended/standard location ($p < .001$), where sounds were most likely to occur. The 0°standard showed slower reaction times at the ±45° locations, but faster reaction times to sounds occurring at the ±90° locations ($p < .001$). Participants responded accurately to the AM-rate on over 95% of trials.

These results demonstrate that attentional bias does not decrease linearly as distance from the attended location increases. First, a sharp decrease in attentional bias occurs at



(a) Single Attended Location **(b) Multiple Attended Locations**

Figure 4: Variables and constraints representing the three maps and their interconnections when the model includes (a) a single attended location or (b) multiple attended locations. For clarity, only the constraints relative to the attended location(s) and the variable corresponding to the [-90°,-88°] locations are shown.

areas surrounding the attended location. Research into visual-spatial attention has produced similar findings. When fixating visual attention, the brain has been shown to actively inhibit visual-attention at areas immediately surrounding the attended location, in order to help isolate attention to that location - termed center-surround inhibition [Itti *et al.*, 1998]. Second, reaction times speed up again for sounds occurring at locations far from the attended location (+/-90°in Figure 3). The auditory system is able to detect sounds coming from any direction, making it excellent at detecting potential oncoming threats [Scharf, 1998]. This on-line threat detection system is believed to be responsible for the increased attentional bias observed at locations far from the attended location, and is accordingly modeled by the saliency map in our model.

3 Extending the model to multiple locations

We now relax the assumption that a sound is expected from a single location, allowing for the case where a sound is equally probable from multiple locations. The constraint-based representation of this more general model remains similar to the previous model described above, except now the model takes k attended locations as input for the goal and saliency map.

Our cognitive hypothesis is that a k-location setting can be modeled as a combination of k copies of the original modeled each centered at one of the locations.

In the new model, the single attended location variable A is now replaced by a set of k location variables $A_j = a_j$, where $j = 1..k$. Each binary constraint (involving only two variables) in the original model, which involved A and each of the V_G^i is now replaced by a constraint involving $k+1$ variables, namely the k locations variables and V_G^i:

Average of k Standard Gaussian Distributions:

$$(A_1 = a_1, ..., A_k = a_k, V_G^i = \frac{G}{k}\sum_{j=1}^{k} e^{\frac{-|a_j - i|^2}{2*d_G^2}}) \qquad (4)$$

As it can be seen the overall bias of the goal map is defined as the average of k identical Gaussians each having its peak at one of the k locations.

Similarly, the k+1-ary constraints below replace the binary constraints of the saliency map in the original model.

Average of k Inverted Gaussian distributions:

$$(A_1 = a_1, ..., A_k = a_k, V_S^i = \frac{G_S}{k} - \frac{G_S}{k}\sum_{j=1}^{k} e^{\frac{-|a_j - i|^2}{2*d_S^2}}) \quad (5)$$

The bias of the saliency map is thus defined as the average of k inverted Gaussians centered at the k locations. The priority map does not change, and remains defined as in Equation 3.

In Figure 4 we depict (partially) in (a) the constraint graph of the original model and in (b) the graph of the new model. In both graphs we have the location variables (green nodes), the goal map variables (red nodes), the saliency map variables (blue nodes) and the priority map variables (purple nodes). Above the nodes corresponding to goal and priority map and under those of the saliency map we depict a graph showing the bias. This should be interpreted as follows: each node corresponds to a variable modeling the bias at a particular location. Such a location is the value on the x-axis. The corresponding value on y-axis is the level of attentional bias, that is the value assigned to the variable corresponding to that node. For example, the left most node of the goal map, represent the variable corresponding to location [-90°,-88°] which has value 0. Constraints are depicted as (hyper)-edges in Figure 4 connecting the nodes corresponding to constrained variables.

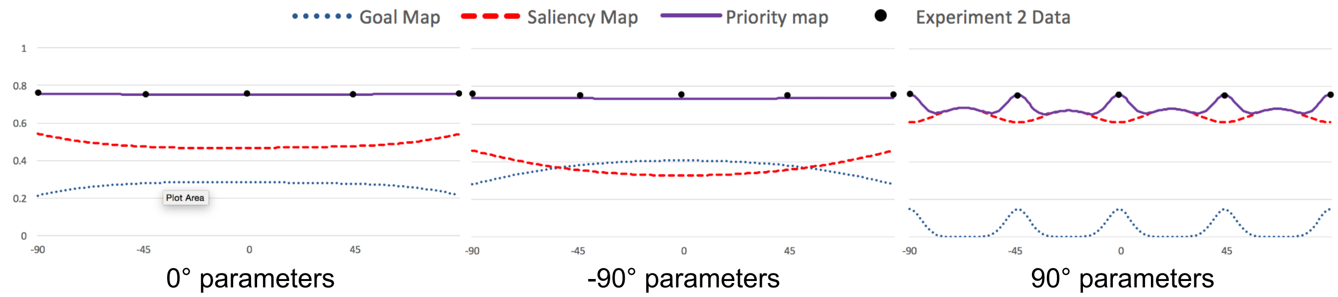


Figure 5: Graphs of the goal, saliency and priority maps in the general model when calculated using parameters found to best fit data at 0°, -90°, and 90° standard locations.

4 Behavioral task and results with multiple attended locations

In this section we describe how the behavioral data for equiprobable attended locations was generated and we discuss how the new model, with the appropriate parameters was able to predict such results.

4.1 Behavioral task

We validate our extended model on data generated by an experiment which utilizes the same approach described in 2.4, but differs in that stimuli were equally likely to occur from all five spatial locations (p=0.2). Accordingly, participants were not instructed to attend to any one particular location during the task. We previously performed an experiment with equal probability at all locations [Golob *et al.*, 2016], but this was done uniquely as a control to ensure the observed attentional distribution was caused specifically by subjects' expectancy at the 'standard' location. The new experiment reproduced these control results in a new group of participants.

As shown in Figure 3 (B) and (C) (dotted line), and in line with previous control results, the results of this experiment demonstrate no significant differences in reaction time across all five locations. These results show that participants tend to distribute attention evenly across all five locations when they are not provided with a standard location. The observed flattening of the curve provides a challenging but potentially fruitful base-case that ultimately let to extending the model.

5 Results

As we have mentioned in Section 2.3, with our original model we targeted the case where sounds are expected from a standard location using Equations 1, 2 and 3 [Golob *et al.*, 2016]. In this model, the peak of attentional bias is centered at the standard location. Stochastic local search was used to fit the parameter values d_G, d_S, G_G, and G_S to the behavioral data obtained for each of three standard locations (0°, -90° and 90°). We used the sum of squared errors to evaluate the fit after each iteration of local search until convergence. The values for the parameters corresponding to the best fits are listed in Table 1. Recall that in the new model, the priority map is the summation of a goal map (created from the average of k gaussians) and a saliency map (created from the average of k inverted gaussians). In the goal map, each of the five gaussians is centered around one of the equally probable locations. Likewise, each of the five minimums of the inverted gaussians are placed at one of these same locations. The key point here is: what are the right parameters for the Gaussians? We conjectured that the parameters learned from the single-location would be a good choice.

To test our hypothesis we plugged in our general model (Equations 4, 5) the parameters found in Table 1 for each of the five Gaussians in the goal and saliency map. We tried this with the parameters learned for the standard 0°, -90° and 90° locations. The resulting goal, saliency and priority maps can be seen in Figure 5. By calculating the sum of squared errors of the new model against the behavioral data (where sounds are equally probable from five locations), we get the fit values found in the first row of Table 2. The fit values indicate relatively small errors, with the best fit model being the one using parameters learned from the +90° data.

To further support our hypothesis that single-location parameters are good predictors for equiprobable-location settings, we ran stochastic local search to find parameter values that could improve the fit further. The local search algorithm was initialized with the single-source parameters and then iterated until convergence (approximately, 1000 iterations), searching for parameters that increased the fit of the behavioral data. The final fit values are indicated in second row of Table 2. From this we can see that the fit improved only very modestly and only when starting at the the $\pm 90°$ parameters.

Std Loc:	d_G	d_S	G_G	G_S	Fit
0°	5.95	15.169	0.764	0.742	0.0009
-90°	48.496	50.989	0.755	0.746	0.0033
90°	4.96	13.87	0.764	0.741	0.0034

Table 1: Parameter values obtained from using a stochastic local search to fit the priority map to the experimental data.

6 Conclusions and Future Work

We have presented a new model of spatial auditory attention that handles a task where sounds come from equiprobable locations. We hypothesized that parameters previously learned from a single-location task would be good predictors

Model	0°	-90°	90°
Before local search	0.004	0.02	0.003
After local search	0.004	0.002	0.002

Table 2: Fit values calculated as the sum of squared error between model results and behavioral data.

for an equiprobable location task. By using these parameters, we were able to achieve good fit with relatively small error against new behavioral data from a task where sounds come from five equally probable locations.

We will further investigate the role of probability in attention gradients using experimental data where sounds are presented away from the standard attended location with 0.04 and 0.12 probability. For the computational model, this will involve handling sequences of stimuli. We will also examine the addition of short-term memory load into the behavioral task, such as when memorizing three words and then performing several trials of the task. Since the load and task-specific information both rely on short-term memory, this should impair top-down control and the goal map. We hypothesize that this should increase reaction times to sounds near the attended location, but not far where the saliency map has a larger influence. Finally, we plan to investigate whether changes in sound intensity decrease reaction times near the standard location, which would be represented by changes to the saliency map. A long-term goal is the embedding of our model as an ACT-R module for auditory spatial attention.

Acknowledgments

This work is supported by NIH under grant number R01-DC015736.

References

[Allport, 1989] Allen Allport. Foundations of Cognitive Science. pages 631–682. MIT Press, Cambridge, MA, USA, 1989.

[Baddeley, 2010] Alan Baddeley. Working memory. *Current Biology*, 20(4):R136–R140, February 2010.

[Caparos and Linnell, 2010] Serge Caparos and Karina J. Linnell. The spatial focus of attention is controlled at perceptual and cognitive levels. *Journal of Experimental Psychology. Human Perception and Performance*, 36(5):1080–1107, October 2010.

[Cave and Bichot, 1999] K. R. Cave and N. P. Bichot. Visuospatial attention: beyond a spotlight model. *Psychonomic Bulletin & Review*, 6(2):204–223, June 1999.

[Cohen *et al.*, 2015] Michael Cohen, Julián Villegas, and Woodrow Barfield. Special issue on spatial sound in virtual, augmented, and mixed-reality environments. *Virtual Reality*, 19(3-4):147–148, 2015.

[Cowan, 1988] N. Cowan. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104(2):163–191, September 1988.

[Dehais *et al.*, 2014] Frédéric Dehais, Mickaël Causse, François Vachon, Nicolas Régis, Eric Menant, and Sébastien Tremblay. Failure to detect critical auditory alerts in the cockpit: evidence for inattentional deafness. *Human factors*, 56(4):631–644, 2014.

[Edworthy *et al.*, 2017] Judy Edworthy, Scott Reid, Siné McDougall, Jonathan Edworthy, Stephanie Hall, Danielle Bennett, James Khan, and Ellen Pye. The recognizability and localizability of auditory alarms: Setting global medical device standards. *Human Factors*, 2017.

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

[Golob *et al.*, 2016] Edward J. Golob, K. Brent Venable, Maxwell T. Anderson, Jesse A. Benzell, and Jaelle Scheuerman. Modelling auditory spatial attention with soft constraints. In *4th International Workshop on Artificial Intelligence and Cognition*, New York City, New York, 2016.

[Golob *et al.*, 2017] Edward J. Golob, K. Brent Venable, Jaelle Scheuerman, and Maxwell T. Anderson. Computational modeling of auditory spatial attention. In *Cog Sci 2017*, London, UK, July 26-29 2017.

[Greenwood and Parasuraman, 1999] P. M. Greenwood and R. Parasuraman. Scale of attentional focus in visual search. *Perception & Psychophysics*, 61(5), July 1999.

[Itti and Koch, 2001] L. Itti and C. Koch. Computational modelling of visual attention. *Nature Reviews. Neuroscience*, 2(3):194–203, March 2001.

[Itti *et al.*, 1998] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.

[Lewandowsky and Farrell, 2010] Stephan Lewandowsky and Simon Farrell. *Computational Modeling in Cognition: Principles and Practice*. SAGE Publications, November 2010.

[Mller *et al.*, 2005] Notger G. Mller, Maas Mollenhauer, Alexander Rsler, and Andreas Kleinschmidt. The attentional field has a Mexican hat distribution. *Vision Research*, 45(9):1129–1137, April 2005.

[Pillsbury, 1908] W. B. Pillsbury. *Attention*. Half-title: Library of philosophy. Ed. by J. H. Muirhead. S. Sonnenschein & Co., ltd. The Macmillan co., London, New York, 1908.

[Rorden and Driver, 2001] C. Rorden and J. Driver. Spatial deployment of attention within and across hemifields in an auditory task. *Experimental Brain Research*, 137(3-4):487–496, April 2001.

[Rossi *et al.*, 2006] Francesca Rossi, Peter van Beek, and Toby Walsh, editors. *Handbook of Constraint Programming*. Elsevier Science Inc., New York, NY, USA, 2006.

[Scharf, 1998] Bertram Scharf. Auditory attention: The psychoacoustical approach. *Attention*, pages 75–117, 1998.

[Zatorre *et al.*, 1999] R. J. Zatorre, T. A. Mondor, and A. C. Evans. Auditory attention to space and frequency activates similar cerebral systems. *NeuroImage*, 10(5):544–554, November 1999.