Computational modeling of auditory spatial attention

Edward J. Golob (Edward.Golob@utsa.edu)
Department of Psychology, University of Texas, San Antonio
San Antonio, TX USA

K. Brent Venable (kvenabl@tulane.edu)
Department of Computer Science, Tulane University
New Orleans, LA USA

Jaelle Scheuerman (jscheuer@tulane.edu)
Department of Computer Science, Tulane University
New Orleans, LA USA

Maxwell T. Anderson (manders7@tulane.edu)
Department of Psychology, Tulane University
New Orleans, LA USA

Abstract

Attention plays a fundamental role in higher-level cognition. In this paper we develop a computational model for how auditory spatial attention is distributed in space. Our model builds on the assumption that attentional bias has bottom-up and top-down components. We represent each component and their synthesis as a map, associating a level of attentional bias to locations in space. The maps and their interaction are modeled using an artificial intelligence approach based on constraints. We describe the behavioral task we have designed to measure the attentional bias and discuss the results. We then test different hypotheses on the shape and interaction modalities of the maps in terms of how well they fit our behavioral data. The findings showed that combining top-down and bottom-up spatial attention gradients that differ in their spatial properties produced the best fit to behavioral data, and suggested several novel mechanisms for future testing.

Keywords: auditory attention; computational modeling; saliency map; constraints.

Cognitive engineering problems and attention

Humans evolved in a dynamic environment of shifting opportunities and threats. Consequently, we are well-equipped to organize and frequently change goals and priorities to effectively deal with events in the natural and social worlds. High-level cognitive attributes, such as intelligence, creativity, and imagination presumably evolved to capitalize on these dynamics to promote survival (Flinn, Geary, & Ward, 2005). A key aspect of higher-level cognition is attention. An important role for attention-like selection in information processing may not be limited to human cognition. For example, Helgason and colleagues proposed that attention is an essential element for systems to exhibit generalized intelligence, regardless of whether it is a biological or artificial intelligence system (Helgason, Thorisson, Garrett, & Nivel, 2014).

In this article we broadly consider attention as a flexible means of enhancing specific aspects of information processing, as determined by factors such as the current goal (top-down) or stimulus characteristics important to the organism (such as unexpected loud sounds)(Chun, Golomb, & Turk-Browne, 2011). This flexibility is assumed to be implemented by specific cognitive processing routines that were selected during the course of human evolution (Cosmides & Tooby, 2013). Differences between sensory modalities in terms of how the adequate stimulus and receptor transduction relate to the kinds of information that can be detected in the environment is one factor relevant to the design of attentional processes. Consequently, in at least some respects attentional processing may sharply differ between sensory modalities.

We focus on the auditory system, and consider implications of the idea that the auditory system has a comparative advantage over other modalities in the ability to panoramically monitor the environment. Hearing provides an early warning system (Scharf, 1998) that allows organisms to prepare for, or evade, threats and to capitalize on prey or mating opportunities. This "3-D sphere" of spatial sensitivity for hearing is unique among sensory modalities because it can detect environmental events that are at a distance from the body (cf. somatosensation, gustation, to some degree olfaction) and out of sight (vision).

Stability-flexibility dilemma and attentional systems

Most attention models consider attention that is directed by a conscious choice ("top-down" or "voluntary") to differ in important ways from attention that is involuntarily "captured" by an event in the world that has a salient property (Petersen & Posner, 2012). Salience can be due to physical properties, such as a loud sound, or by having personal meaning such as one's name (Moray, 1959) or taboo words (Arnell, Killman, & Fijavz, 2007),
and other aspects that may depend on the situation (Gygi & Shafiro, 2011). The distinction between top-down and bottom-up attentional functions is both useful and

meaningful, even though top-down and bottom-up attentional processes are highly interactive (Folk, Remington, & Johnston, 1992).

One of the defining features of the top-down aspect of attention is that it is limited. Either by design, such as matching the limitations in the number of actions that can be done at one time, or by overload from having too much information to be processed at one time, or both, voluntary attention is limited (Allport, 1989; Posner, 1978). Spatial attention has been intensively studied, in part, because it vividly illustrates limitations in attentional capacity. The limited capacity of spatial attention can be expressed as a spatial gradient relative to an attended location (reviewed in (Cave, 2013). The classic way to consider this gradient is that it reflects decreased investment of attentional resources with greater distance from an attentional focus, and the extent of the gradient can be adjusted based on the current task (Eriksen & St. James, 1986).

The fundamental problem with including top-down and bottom-up attention in one general attention system that distributes attentional resources across space is that attention cannot be simultaneously both focused and diffuse. This kind of trade-off has been termed the "stability-flexibility dilemma (Liljenström, 2003), the "shielding–shifting dilemma" (Thomas Goschke & Bolte, 2014), and a trade-off between organization and flexibility (Baars, 1997). The problem is compounded by not knowing when something will happen outside of the attentional focus that is critical for survival, thus preventing an anticipatory shift by top-down attention. Both top-down and bottom-up attention have clear survival value, but limited attention capacity implies trade-offs between resources devoted to top-down vs. bottom-up attention functions. Similar issues concerning cognitive trade-offs have been explored in the context of cognitive control and task switching (Goschke, 2000), automatic vs. controlled processing (Schneider & Chein, 2003), various dual process models of cognition (Evans, 2008), long-term knowledge (Caramazza & Shelton, 1998), and memory systems (Sherry & Schacter, 1987).

Methods and computational modeling

The present study addresses the stability-flexibility dilemma posed by needing attention to be both focused on a task while also monitoring the environment for potential threats or opportunities by modeling auditory spatial attention bias as the net result of two attention modules and their output (Figure 1). Our aim is to develop a rigorous quantitative theory of auditory spatial attention. One module, called the "goal map" is devoted to top-down attention necessary to perform the current task. The other module, termed the "saliency map", is specialized to monitor, in parallel, the environment and, when needed, engage bottom-up orienting that overrides current attentional focus based on top-down processes. We combine novel parametric behavioral measures to map-out auditory attention over space with a computational model to explain how specific top-down and bottom-up mechanisms jointly determine the shape of auditory spatial attention gradients.

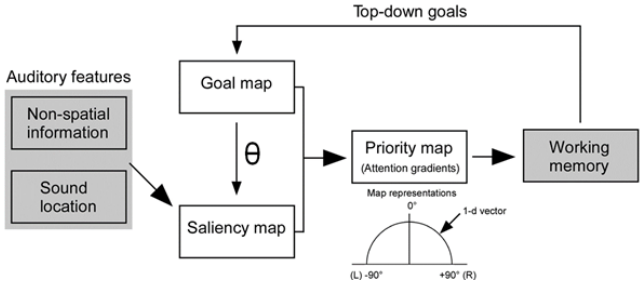

Figure 1. Proposed attentional model architecture.

Relative to existing models of auditory attention, the current model is designed to help understand somewhat higher levels of cognitive processing. Others have modeled perceptual features and how they are combined to generate a saliency map. There is overlap with our model at the level of saliency map. Prior work computes perceptual features such as stimulus location, frequency, intensity, and saliency as an output that is computed from a raw sensory input (Coensel & Botteldooren, 2010; Kayser, Petkov, Lippert, & Logothetis, 2005). Instead, we start with perceptual features as a given input and model how top-down and bottom-up modules interact in the context of working memory. Note also that the choice of modeling auditory spatial attention in the frontal plane has the benefit of needing to explain attentional bias in only one-dimension (the azimuth plane at a constant distance from center of head), which simplifies modeling. In contrast, visual studies of saliency maps use two-dimensional models (Kalinli & Narayanan, 2007).

Model design The model is designed using constraints, a very general and powerful artificial intelligence framework for problem modeling and solving. (Rossi, Van Beek, & Walsh, 2006). Constraints lie at the core of many successful applications in several domains such as scheduling, planning, vehicle routing, configuration, networks, and bioinformatics. The basic idea in constraint-based modeling 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, if it exists, an assignment to all the variables that satisfies all the constraints. A constraint concerns a subset of variables and defines which simultaneous assignments to those variables are allowed. Solutions are found by searching the solution space either systematically, as with backtracking or branch and bound 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. Constraints have been used before in the context of human cognition for example to model skilled behavior (Howes, Vera, Lewis, & McCurdy, 2004). Recently an implementation of the cognitive architecture ACT-R

based on constraint handling rules, which are a closely related to constraints, has been proposed in (Gall & Frühwirth, 2014). To the best of our knowledge, this is the first time constraints are employed at this level of cognitive modeling and in the context of attention.

The model makes several assumptions regarding proactive and reactive control. According to Braver's dual mechanisms framework (Braver, 2012), proactive control generates a sustained attentional bias in accordance with task goals, such as focusing on a pianist about to begin their recital. Reactive control, as the name suggests, is attentional orienting in response to a stimulus, such as if the pianist plays their first chord and everybody realizes that the piano is out of tune. In our model the goal map is the mechanism for proactive control. The spatial focus of the goal map can also be redirected in response to stimuli, and so could have a role in reactive control too. In contrast, the saliency map codes for reactive control. The relation between the saliency map and proactive control is only indirect. The focus of the saliency map is designed to be away from the goal map focus, thus any proactive shifts in the goal map focus will consequently lead to a similar shift in the saliency map focus.

Task and data to be modeled Young adult subjects (n=42) listened to 25 and 75 Hz amplitude modulated white noise, and responded with left/right hand (counterbalanced across subjects). Virtual stimuli were delivered via headphones to one of 5 locations in the frontal horizontal plane (L→ R locations: -90°, -45°, 0°, +45°, +90°; 2.4 sec SOA). In each 6 min block subjects attended to a standard location (either -90°, 0°, or +90°). Most stimuli were given at the standard location (p=.84), with occasional shifts to the other 4 locations (p=.04/location). Analysis of variance (ANOVA) was used to examine reaction time as a function of standard condition (3) and stimulus location (5). Data were collapsed across AM rates (ns). We note that the following model is designed from general principles based on the attention and working memory literature, but the actual modeling here is very specific to our task. This is common in other areas such models of canonical visual search tasks. Future work will expand this model to include other tasks and situations.

The model Behavioral results were modeled using a constraint-based representation made up of three components: goal map, saliency map, and priority map. The maps represent the attentional bias across the horizontal frontal plane (-90° to +90°) (see heat-map in Figure 2, top-left) . The priority map is the weighted sum of the goal and saliency maps and represents the total attentional bias at each degree location. Operationally, attention bias in the priority map relates to reaction time by equation 1:

Eq. 1 Attentional bias = (2,000-reaction time)/(2,000)

The "2,000" value was chosen as an upper limit on reaction times to be analyzed (both in ms), and included nearly every correct trial in every subject. The units of attention bias are arbitrary, but index reaction time with a range of between approximately 0.90, which corresponds to a an extremely fast reaction time of 200 ms, to 0.0, which indicates a 2,000 ms reaction time. Thus, larger attention bias values in the priority map reflect short reaction times and efficient processing, and longer reaction times have smaller values.

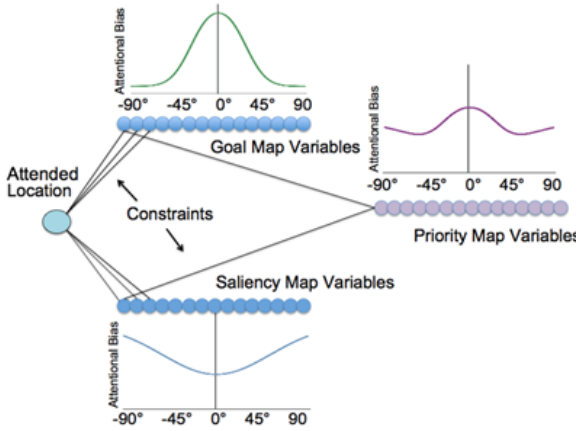

Figure 2: Variables and constraints that represent the interactions between the three maps in the model.

Each map is represented by a collection of variables, one for each 2-degree (the minimum distinguishable by a human ear) location in the range {-90°,…,+90°}, and a set of constraints over the variables. Figure 2 (left) shows a portion of the constraint graph of the model where nodes correspond to variables and edges to constraints. These constraints limit the simultaneous assignments of the constrained variables as indicated in the equations below, where V_G^i, V_S^i, and V_P^i represent the *i*-th variables of the goal, saliency and priority maps. The constraints defining each map involve the variable corresponding to the attended location (*A*) and the variables corresponding to a location. The variables associated with the goal map (blue nodes in Figure 2) are constrained to represent a standard Gaussian distribution with its peak at level G_G and located at the attended location A. An example of one such distribution is shown in Figure 2 right above the set nodes representing the goal map variables. Each node represents the 2 degree portion of the x-axis right above it and the associated attentional bias value (y-axis) is the value assigned to the variable corresponding to the node. Similarly, for the saliency (where the distribution is shown below the nodes) and the priority map. Note that parameter d_G is the standard deviation of the Gaussian and is used to model a symmetrical decrease in top-down attentional resources away from the goal location. Likewise, the variables corresponding to the saliency map have values compatible with an inverted Gaussian distribution with peak level $-G_S$ at attended location A and standard deviation d_S representing a symmetrical increase in bottom-up attention away from the attended location. Finally, each priority map's variable takes as value the weighted sum of the values of the corresponding goal map and saliency map variables. The

graph at the bottom-right of Figure 2, shows an example of a goal map bias (green), saliency map bias (blue) and the associated priority map (purple). Weights α and β, are used to model the magnitude of contribution of, respectively, the goal and saliency maps to the priority map.

$$Goal\ Map: (A = a, V_G^i = G_G e^{\frac{-|a-i|^2}{2d_G^2}})$$

$$Saliency\ Map: (A = a, V_S^i = G_S - G_S e^{\frac{-|a-i|^2}{2d_G^2}})$$

$$Priority\ Map: (V_G^i = u, V_S^i = v, V_P^i = \alpha u + \beta v)$$

We note that the constraint based model allows an easy extension to a 2D or 3D bias distribution. This can be achieved in two ways: either by increasing the number of variables (e.g. for the planar case a set for each concentric hemisphere), or by increasing the complexity of the domain values, e.g. 2D bias distributions over 2 degree sectors.

Results

We first describe and discuss the results of the behavioral experiment in the previous section, and then the results of applying our computational model to the behavioral data. The goal map and saliency map each had two parameters for fitting: attention bias and standard deviation.

Behavioral results

Reaction time curves for angular shifts had an inverted-u shape at all 3 standard locations (p's<.01)(Figure 3). Attending to the right (+90° standard) had an attenuated inverted-u curve vs. -90° and 0° standards (p<.01). Results show comparable reaction time increases to the nearest shift location for the 0° and -90° standards, and then decreases in reaction time at the most distant locations. For the +90° standard there was a more gradual increase and decrease in reaction time across shift locations. Accuracy was high for all stimulus locations and conditions (> 94%) and will not be analyzed here.

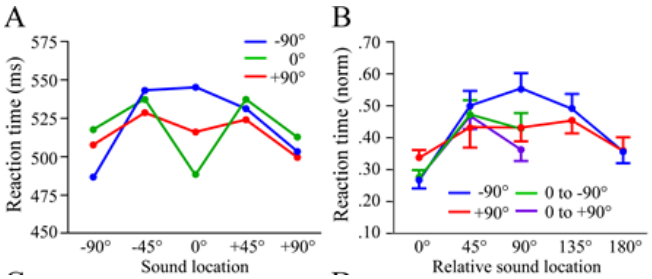

Figure 3. Behavioral results showing mean reaction times for standard locations at the far left (-90°), midline (0°), and far right (+90°) locations. A) Reaction time as a function of location for the three standard locations. B) Normalized reaction times plotted relative to the standard location, here denoted by "0°".

Computational modeling results

Stochastic local search was used to find parameter values for d_S, d_G, G_S and G_G, that minimize the root-mean-square (rms) error between the priority map and behavioral data. Bootstrapping methods were used to compare model fit as the parameters for the goal and saliency maps varied. There were 100 runs for each standard location to assess the consistency of results. On each run half of the subjects (n=21) were randomly selected to train the model. The model was then tested for fit using root-mean square error on the grand average of the remaining subjects (n=21).

Comparison of two vs. three-component models Having attention bias centered on the standard location and decreasing with distance was modeled with only the goal map having input to the priority map. This two-component model had a poor fit to the reaction time data, with rms values nearly 100x worse than models with both goal and saliency map inputs to the priority map (Figure 4). Models with both top-down (goal map) and bottom-up (saliency map) spatial attention bias fit the data well, with rms values ranging from 0.0040 to 0.0035 for left or right standard locations (±90°) and 0.0011 and 0.0012 for the 0° standard. The fits at each standard location were all significantly different from each other (p's < .001). By contrast, rms values with only the goal map in the model were 0.3137 ((-90° standard), 0.3060 (+90° standard), and 0.1191 (0° standard). We note that a model based only on the saliency map was not tested as it would have not been able to model the increased bias at the attended location. The results clearly show that a simple attention gradient that decreases with distance from the attended standard location (goal map only) is unable to account for the behavioral data. Models with both goal and saliency maps provided a good fit to the behavioral results. It is unclear why

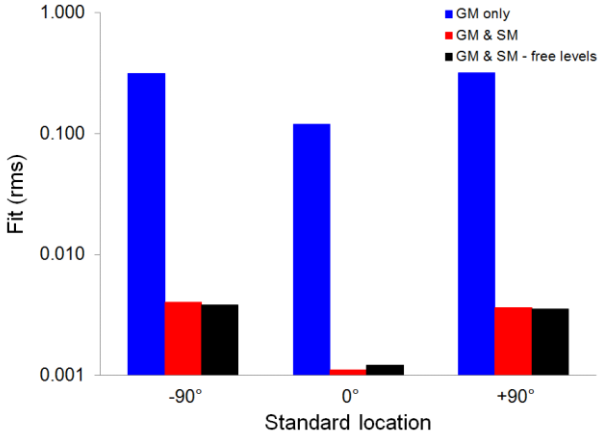

Figure 4. Model fit with only the goal map (GM) representing top-down attention bias vs. the addition of a bottom-up component (saliency map, SM). Models examined whether goal and saliency maps had either equal influence on the priority map ("GM & SM) or their levels were included as a parameter in the model ("free levels"). Model fit was measured using root-mean-square error (rms).

the fit for the 0° standard is even better than the ±90° standards, but this may relate to the truncated range of locations on either side (±90°).

Standard deviation parameters The range of spatial attention bias for the goal and saliency maps was quantified with separate standard deviation parameters (Figure 5). When only the goal map was included in the model the best fits had standard deviations of ~100°, which produced a gradual decrease of attentional bias from the standard location. As shown above, only including the goal map produced a poor fit to the behavioral data. In all models with goal and saliency maps the standard deviations had, large, progressive reductions from standard locations on the left, to midline, and to the right ($p<.001$). This pattern was evident for both the goal and saliency map SD parameters. Analysis of both fixed and free bias models showed main effects of map type, with significantly larger SD values in the saliency map (p's $< .001$). There were interactions between standard location and map type, indicating that the difference between the SD of goal and saliency maps varied among standard locations (p's $< .001$).

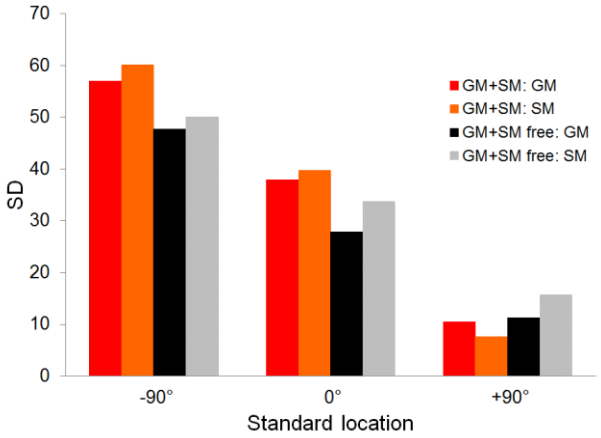

Figure 5. Standard deviation (SD) parameters in the models with goal map (GM) and saliency map (SM) components. Standard deviation units are in degrees.

Attention bias level Lastly, we tested a model where the attention bias levels from the goal and saliency maps to the priority map were free to vary. The findings from when bias parameters were added to the model are shown in Figure 6 for each standard location. For the ±90° standards the goal map had a significantly greater bias than the saliency map, indicating a greater influence over the priority map outcomes. This was most evident for the -90° standard, which had little variability among modeling runs ($p<.001$). In contrast, for the 0° standard there was substantial variability over modeling runs, and there was no significant difference between goal and saliency map bias.

Note that the range of attentional bias levels in the goal and saliency maps is much larger than the priority map (data not shown). This is the result of the model solutions having SD

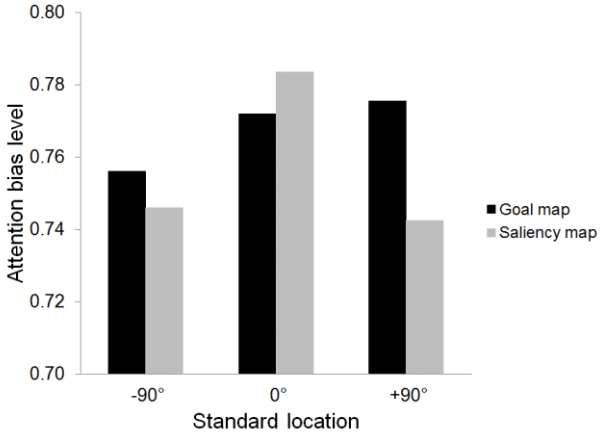

Figure 6. Attention bias level results in the free level models. Bias indicates the overall level of inputs from the goal and saliency maps to the priority map. Greater bias indicates more influence over the priority map values.

parameter values that were both narrow enough to individually have bias levels near asymptote within the degree range tested. The model sums the contributions of goal and saliency maps to generate the priority map, which in turn is proportional to reaction time. The goal and saliency curves over space overlapped such that when one map had low bias the other had a large amount of bias. This additive approach in combination with moderate SD ranges forces many locations to have large differences between goal and saliency map values while retaining a much smaller range of priority map values. For perspective, the range of biases of between .76 - .70 corresponds to reaction times between 480-600 ms.

Discussion and conclusions

In this paper we have studied spatial attention of the auditory system from a behavioral and computational modeling point of view. The main findings were that a traditional top-down attention gradient could not account for the behavioral data, but a model with two gradients corresponding to top-down and bottom-up bias worked well. The model is based on structuring the overall allocation of attentional bias as the sum of bottom-up and a top-down components. We have presented behavioral results aimed at describing the effect of the overall attentional bias and we have provided an experimental evaluation of different model hypothesis in terms of how well they fit the data. There was a pronounced left-right asymmetry in the reaction time profiles as a function of location that was accounted for by progressive reductions in the SD parameters of goal and saliency maps. The results support our approach which constitutes, to the best our knowledge, the first computational model that integrates top-down and bottom-up auditory spatial attention processes.

Acknowledgments

This study was supported by NIH grant DC014736.

References

Allport, A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of Cognitive Science* (pp. 631–682). Cambridge, MA: MIT Press.

Arnell, K. M., Killman, K. V, & Fijavz, D. (2007). Blinded by emotion: target misses follow attention capture by arousing distractors in RSVP. *Emotion (Washington, D.C.)*, *7*(3), 465–477.

Baars, B. J. (1997). The Global Workspace Theory of Consciousness. *Journal of Consciousness Studies Studies*, *4*(4), 292–309.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–13.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1–34.

Cave, K. R. (2013). Spatial attention. In *The Oxford Handbook of Cognitive Psychology* (pp. 117–130). New York, NY: Oxford University Press.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annu Rev Psychol*, *62*, 73–101.

Coensel, B. De, & Botteldooren, D. (2010). A model of saliency-based auditory attention to environmental sound. *Proc 20th Intl Cong Acoustics*, *20*(August), 1–8.

Cosmides, L., & Tooby, J. (2013). Evolutionary Psychology: New Perspectives on Cognition and Motivation. *Annu. Rev. Psychol*, *64*, 201–29.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Percept Psychophys*, *40*, 225–240.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.

Flinn, M. V., Geary, D. C., & Ward, C. V. (2005). Ecological dominance, social competition, and coalitionary arms races: Why humans evolved extraordinary intelligence. *Evolution and Human Behavior*, *26*(1), 10–46.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *J Exp Psychol Hum Percept Perform*, *18*(4), 1030–1044.

Gall, D., & Frühwirth, T. (2014). A formal semantics for the cognitive architecture ACT-R. In *International Symposium on Logic-Based Program Synthesis and Transformation* (pp. 1–18).

Goschke, T. (2000). Intentional Reconfiguration and Involuntary Persistence in Task Set Switching. Control of cognitive processes:, 18, 331. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII* (pp. 331–355). Cambridge, MA: MIT Press.

Goschke, T., & Bolte, A. (2014). Emotional modulation of control dilemmas: the role of positive affect, reward, and dopamine in cognitive stability and flexibility. *Neuropsychologia*, *62*, 403–423.

Gygi, B., & Shafiro, V. (2011). The incongruency advantage for environmental sounds presented in natural auditory scenes. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(2), 551–65.

Helgason, H. P., Thorisson, K. R., Garrett, D., & Nivel, E. (2014). Towards a General Attention Mechanism for Embedded Intelligent Systems. *International Journal of Computer Science and Artificial Intelligence*, *4*, 1–7.

Howes, A. H., Vera, A., Lewis, R. L., & McCurdy, M. (2004). Cognitive Constraint Modeling : A Formal Approach to Supporting Reasoning About Behavior Cognitive Constraint Modeling : A Formal Approach to Supporting Reasoning About Behavior. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 595–600).

Kayser, C., Petkov, C. I., Lippert, M., & Logothetis, N. K. (2005). Mechanisms for allocating auditory attention: an auditory saliency map. *Current Biology : CB*, *15*(21), 1943–1947.

Liljenström, H. (2003). Neural stability and flexibility: a computational approach. *Neuropsychopharmacology 28 Suppl 1*, S64-73.

Moray, N. (1959). Attention in dichotic listening: affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*, 73–89.

Posner, M. I. (1978). *Chronometric explorations of mind*. New York: Halsted Press.

Rossi, F., Van Beek, P., & Walsh, T. (2006). *Handbook of Constraint Programming (Foundations of Artificial Intelligence)*. Amsterdam: Elsevier.

Scharf, B. (1998). Auditory attention: The psychoacoustical approach. In H. Pashler (Ed.), *Attention* (pp. 75–117). East Sussex, UK: Psychology Press.

Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: Behavior, theory, and biological mechanisms. *Cognitive Science*, *27*(3), 525–559.

Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review*, *94*(4), 439–454.