Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 145 (2018) 797–804

Procedia Computer Science

www.elsevier.com/locate/procedia

Postproceedings of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2018 (Ninth Annual Meeting of the BICA Society)

Modeling spatial auditory attention in ACT-R: a constraint-based approach

Jaelle Scheuerman[a,*], K. Brent Venable[a], Maxwell T. Anderson[b], Edward J. Golob[b]

[a] Tulane University, Department of Computer Science, New Orleans, LA 70118, USA
[b] University of Texas at San Antonio, Department of Psychology, San Antonio, TX 78249, USA

Abstract

Attention is the focus of a considerable amount of research in cognitive modeling. Yet, most of the work has been devoted to studying visual attention. Auditory attention differentiates itself from visual attention in many ways. For example, it is often important for our survival to attend to auditory events at a distance and out of sight. Due to limitations on attentional resources, audition must strike a balance between monitoring the environment and attending to current tasks. In this paper, we provide an overview of our previous work which models auditory attention as a spatial gradient made up of a combination of top-down and bottom-up influences [14, 15]. We also present how the model is currently integrated with the audio module in ACT-R. Our approach uses the well-established AI framework of constraint satisfaction problems to model how auditory attention is allocated over space and it 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 reproducing behavioral data of experiments where there is a single attended location. Its integration into a cognitive architecture opens up new possibilities for evaluation in the context of other cognitive functions and in modeling tasks and designing systems where audition is important.

Keywords: cognitive modeling, auditory attention, constraint satisfaction problems, cognitive architectures

1. Introduction and Motivation

Audition is unique from other senses in that it monitors the environment for sounds happening all around us, including those at a distance, behind us, and out of sight. This makes the auditory system particularly useful as an early warning

* Corresponding author.
E-mail address: jscheuer@tulane.edu

1877-0509
10.1016/j.procs.2018.11.028

system that can orient attention to important events, such as the sound of a snapped twig from a predator hiding in the brush behind us. Though these attentional shifts are important to our survival, the auditory system must often balance between attending to potential threats or opportunities, and focusing on a current task.

While visual attention has attracted a substantial amount of attention from the research community, less work has been devoted to modeling spatial auditory attention. The goal of this project is to better understand spatial auditory attention and how top-down and bottom-up processes work together to govern shifting attention to infrequent distractors. Our previous work has analyzed a sustained spatial auditory attention task that measures reaction times to auditory stimuli presented at different locations in space [14, 15]. Using these results, we proposed a computational model that represents auditory attention as a spatial gradient made up of a combination of top-down and bottom-up attentional biases. This model was successful in modeling the sustained spatial auditory attention task.

Here we present an overview of our computational model of spatial auditory attention and show how it has been integrated into the cognitive architecture, ACT-R. ACT-R provides many modules and tools that support developing models in the context of cognition as a whole [2]. Integrating the model into ACT-R allows us to examine spatial auditory attention with other aspects of cognition, such as short-term memory. We foresee that this work will help expand the modeling capabilities of the ACT-R audio module and advance understanding of basic issues in attention, such as top-down and bottom-up interactions, vigilance and capacity limitations. It will also open 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 [9], or when clinicians are unable to distinguish between auditory alarms in a hospital environment [10]. By examining spatial auditory attention in these contexts, it may be possible to predict when such warnings might be ignored.

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 AI approach we adopt using constraints.

2.1. Spatial 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 [23]. This dichotomy is intuitive and has many names in the literature (e.g. top-down/bottom-up, endogenous/exogenous, controlled/automatic [7]. Here 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 the top-down task set. Although the top-down and bottom-up distinction is meaningful, as a practical matter they are highly interactive [12]. 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. Next, we briefly review work on auditory spatial attention at a cognitive level of analysis, and draw from the visual literature when needed to present major points relevant to auditory spatial attention.

Attention can be expressed as a spatial gradient relative to an attended location [21]. Gradients are presumably a byproduct of limited perceptual input capacity, although limitations in behavioral output may also be relevant [1]. The spatial extent of attentional processing is variable [29], and can be modified by directly cuing different size areas [16], or manipulations of perceptual or memory loads [18]. Splitting attention between locations and multi-object tracking are also possible [3, 6]. The ability to deliver attentional benefits rapidly diminishes over time, a phenomenon called the vigilance decrement [20]. This is important because in everyday life attention is commonly deployed over relatively long time periods (e.g. conversation, listening to music).

Auditory spatial cuing decreases reaction times to subsequent targets at a cued location relative to uncued locations [24, 27, 30, 25]. Both Mondor and Zatorre (1995) and Rorden and Driver (2001) 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, with reaction times increasing and then decreasing away from the cued location [22, 5](Mexican-hat shape). This is similar to our findings in the auditory modality, but the auditory results have a much larger spatial range.

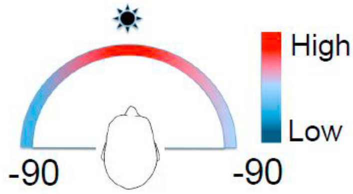

Fig. 1. Example of a gradient of attentional bias from left (-90°) to right (+90°).

2.2. *Constraints and Computational Models*

Constraint programming [26] 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 satisfies the constraints. Constraints concern subsets of variables and define which simultaneous assignments to those variables are allowed. For example, in scheduling activities in a company, the decision variables might be the starting times and the durations of the activities and the resources needed to perform them, and the constraints might be on the availability of the resources and on their use by a limited number of activities at a time.

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 [28] and learning [11]. Recently an implementation of ACT-R based on constraint handling rules, which are a closely related to constraints, has been proposed in [13]. Casting our model into a well-established AI framework will facilitate future generalization of other aspects concerning attention.

3. Behavioral Task

Here we describe a behavioral task designed to map attentional gradients. In the next section, we will describe a computational model and show how it is capable of representing the behavioral data. In the task, white noise is presented from the 5 locations in the frontal plane ($-90°, -45°, 0°, +45°, +90°$), and subjects respond in each trial by discriminating a non-spatial feature (amplitude modulation (AM) rate, 25 or 75 Hz). The slow AM rate sounds like a deck of cards being shuffled while the faster rate is perceived as a buzz. Most stimuli come from a standard location ($p = .84$) but sometimes shift to a distractor location ($p = .04/\text{location}$). Separate blocks have the standard at -90°, 0°, or +90°(counterbalanced).

Figure 2 plots reaction times and location for each standard condition in absolute space (A), as well as the deviant location relative to the standard location (B). There were two main results. First, all conditions had slower responses to distractors vs. standards ($p < .001$), indicating attention shift costs. The reaction time x location function is more prominent for the left vs. the right 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 ±45° locations, similar to the left standard, but a decrease for the ±90° locations, similar to the right standard ($p < .001$). Accuracy was very high (> 95%). The basic results were replicated in new subjects ($n = 12, p < .01$). (C). Second, in each condition reaction times sped-up for the farthest distractor location ($p < .001$). 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 (goal map). Instead, the heightened bias to far distractors is modeled

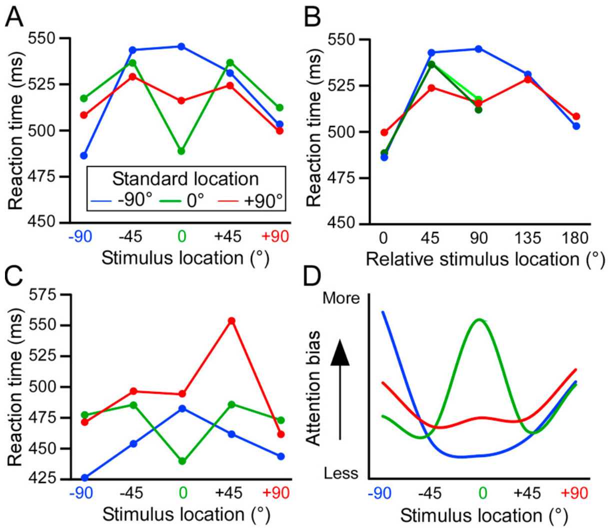

Fig. 2. Reaction time and modeling of basic attention task.

by the saliency map. A control condition in new subjects and equal probability at all locations had no differences in reaction time (n = 20, p = .83), ruling out accounts based on perceptual differences among locations. In D the inverse of the reaction times is given to show how attention bias is theorized to relate to reaction time.

4. Computational Model of Auditory Attention

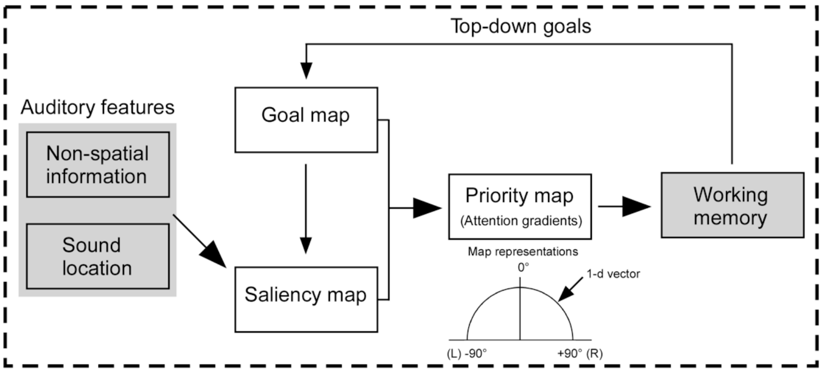

Fig. 3. Computational Model Schematic.

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 [17, 19]. Our model uses basic ideas of top-down and bottom-up attentional control from prominent verbal models [4, 8]. The novelty of the approach we consider here is the application to spatial auditory 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. In previous work, [14, 15] 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 3): 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 1). 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. In the model described in [14], 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:

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

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:

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

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.

$$(V_G^i = u, V_S^i = v, V_P^i = u + v). \tag{3}$$

5. Model Results and Discussion

This model was validated on results obtained from the behavioral task by testing it against different options for the goal and saliency map as described in [14]. The shapes represented by Equations 1 and 2 emerged as the best in terms of fitting the experimental data for all three standard locations.

We evaluated the model by calculating the sum of squared errors, $E(p)$, between the model and the behavioral data:

$$E(p) = \sum_{x \in \{-90°,-45°,0°,+45°,+90°\}} (d_x - p(x))^2$$

where, d_x is the bias associated to location x in the experimental data and $p(x)$ is the value associated to x by the priority map. Stochastic local search was used to fit the model parameters (G_G, d_G, G_S, and d_S) for all three attended locations (-90°, 0°and +90°). The resulting parameters are summarized in Table 1.

The parameters in Table 1 specify the shape of the goal and saliency maps that best combine to create a priority map that fits the behavioral data with relatively low error. The shape of each map compared to the behavioral data can be seen in Figure 5. In all three standard locations, the G_G parameter is slightly higher than the G_S parameter, indicating more attentional bias at the goal location than locations furthest from the location. This is reasonable when compared with our behavioral data which shows fastest reaction times at the goal location, and nearly as fast reaction times at locations from the goal location. The d_G and d_S parameters for the -90°standard location are much larger than the other standard locations, which models the asymmetrical differences seen in the behavioral data.

	G_G	d_G	G_S	d_S	$E(p)$
-90°	0.7463	40.6742	0.7342	42.5800	0.0068
0°	0.7603	6.9700	0.7463	15.8907	0.0044
90°	0.7511	8.9600	0.7325	12.9400	0.0036

Table 1. Model parameters resulting in the best fit at each standard location.

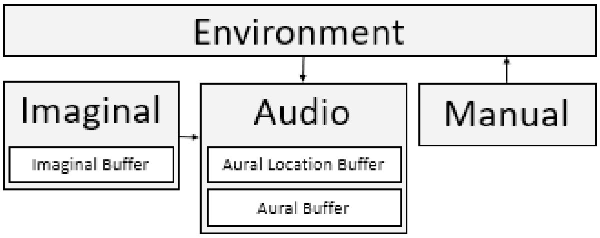

Fig. 4. ACT-R modules used in the modeled task. Sound features in the Imaginal buffer affect attentional bias, increasing or decreasing the time needed to move a sound event from the Aural Location buffer to the Aural Buffer.

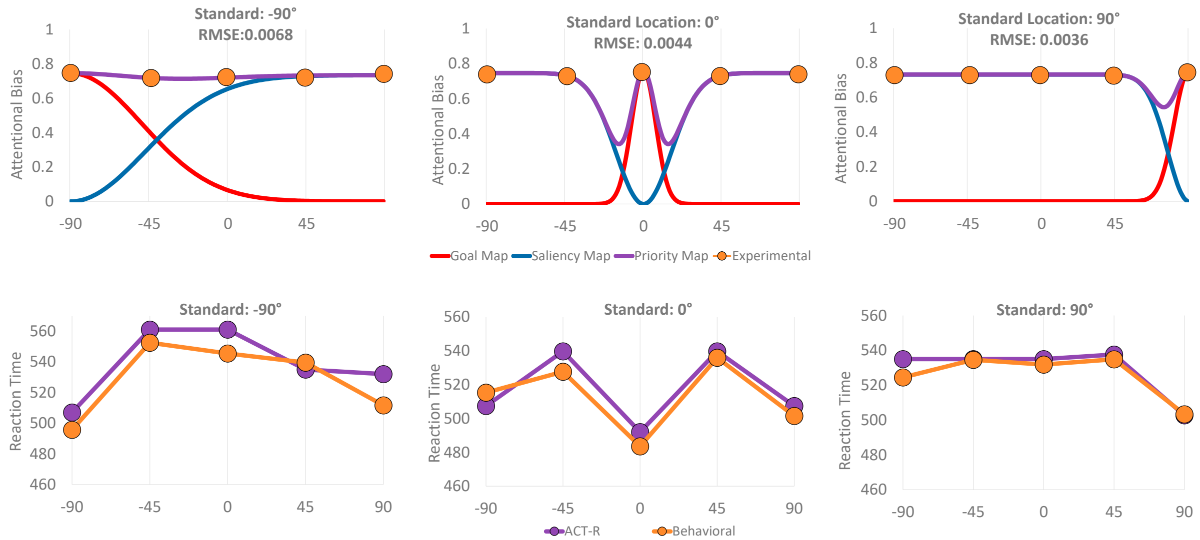

Fig. 5. Top row: Constraint model versus behavioral results. Bottom row: Simulated ACT-R results versus behavioral results.

6. ACT-R Integration

The constraint model provides a first example of how attentional bias can be modeled as a combination of top-down and bottom-up influences and shows how the sound location constrains the shape of the goal and the saliency maps. We hypothesize that there are other influences on the shape of the goal and saliency map, such as retaining words or items in short term memory, the loudness or frequency of the sound, and the likelihood of sound presentation at each location. To investigate how the goal map and saliency map might be affected by other cognitive functions, we integrated the constraint model into ACT-R.

ACT-R provides a basic Audio module that allows cognitive agents to attend and respond to simulated sounds. When a sound becomes available to the agent's environment, it appears as a sound event in the audicon. Sound events include attributes such as the type of sound, its location and content. Once the sound is available in the audicon, the

ACT-R agent can attend to that sound automatically (using buffer stuffing) or with a production rule that instructs the agent to attend to a sound meeting certain constraints. Attending to the sound means first adding it to the Aural-Location buffer, representing a sound that has been perceived, but not encoded. After the sound has been perceived and encoded, it becomes available in the Aural buffer. ACT-R assumes a constant amount of time for placing a sound in the Aural-Location and Aural buffers. This simplified view of auditory attention was not expressive enough to represent the affect of attentional bias and the range of reaction times observed in our behavioral data. Because of this, we extended the Audio module to include a representation of the priority map and calculate response times based on attentional bias.

Like the human subjects in our behavioral task, the ACT-R agent is instructed to attend to a certain location and then press a button depending on the AM-rate of the sound. In ACT-R, this consists of storing the attended location in the Imaginal buffer, which is meant to contain task relevant information. The Motor module handles the appropriate key press when production rules indicate that a sound is in the Aural buffer.

The ACT-R agent we designed includes four production rules to govern its 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".

To incorporate attentional bias into the reaction times, we use the priority map as calculated by our auditory attention module. The location stored in the Imaginal buffer represents the attended location (A). The four parameters of the priority map learned from the behavioral data are provided to the module, including the standard deviation (d_G) and peak value (G_G) of the goal map, and standard deviation (d_S) and minimum value (G_S) of the saliency map. Given these parameters, the attended location and the location of a sound event, the attentional bias is calculated and a response time is returned.

Using this extension to the Audio module, we were able to build a cognitive agent that was 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 ACT-R simulated data compared to the behavioral data can be seen in Figure 5.

7. Future Directions

Integrating our constraint-based model into the ACT-R cognitive architecture opens up new possibilities for modeling tasks that examine how the attentional gradient is affected by other aspects of cognition, such as learning and memory. This research may also lead to identifying additional areas where the ACT-R audio module may be extended to provide a general model of auditory attention and pave the way for modeling tasks and designing systems where audition is important.

8. Acknowledgements

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

References

[1] Allport, A., 1989. Foundations of Cognitive Science, MIT Press, Cambridge, MA, USA, pp. 631–682. URL: http://dl.acm.org/citation.cfm?id=102953.102969.

[2] Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y., 2004. An integrated theory of the mind. Psychological review 111, 1036.

[3] Awh, E., Pashler, H., 2000. Evidence for split attentional foci. Journal of Experimental Psychology. Human Perception and Performance 26, 834–846.
[4] Baddeley, A., 2010. Working memory. Current Biology 20, R136–R140. URL: http://www.sciencedirect.com/science/article/pii/S0960982209021332, doi:10.1016/j.cub.2009.12.014.
[5] Caparos, S., Linnell, K.J., 2010. The spatial focus of attention is controlled at perceptual and cognitive levels. Journal of Experimental Psychology. Human Perception and Performance 36, 1080–1107. doi:10.1037/a0020367.
[6] Cavanagh, P., Alvarez, G.A., 2005. Tracking multiple targets with multifocal attention. Trends in Cognitive Sciences 9, 349–354. doi:10.1016/j.tics.2005.05.009.
[7] Chun, M.M., Golomb, J.D., Turk-Browne, N.B., 2011. A taxonomy of external and internal attention. Annual Review of Psychology 62, 73–101. doi:10.1146/annurev.psych.093008.100427.
[8] Cowan, N., 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. Psychological Bulletin 104, 163–191.
[9] Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., Tremblay, S., 2014. Failure to detect critical auditory alerts in the cockpit: evidence for inattentional deafness. Human factors 56, 631–644.
[10] Edworthy, J., Reid, S., McDougall, S., Edworthy, J., Hall, S., Bennett, D., Khan, J., Pye, E., 2017. The recognizability and localizability of auditory alarms: Setting global medical device standards. Human Factors .
[11] Epstein, S.L., 1994. For the right reasons: The FORR architecture for learning in a skill domain. Cognitive Science 18, 479–511. URL: http://www.sciencedirect.com/science/article/pii/0364021394900175.
[12] Folk, C.L., Remington, R.W., Johnston, J.C., 1992. Involuntary covert orienting is contingent on attentional control settings. Journal of Experimental Psychology. Human Perception and Performance 18, 1030–1044.
[13] Gall, D., Frühwirth, T.W., 2014. Exchanging conflict resolution in an adaptable implementation of ACT-R. TPLP 14, 525–538. URL: http://dx.doi.org/10.1017/S1471068414000180, doi:10.1017/S1471068414000180.
[14] Golob, E.J., Venable, K.B., Anderson, M.T., Benzell, J.A., Scheuerman, J., 2016. Modelling auditory spatial attention with soft constraints, in: 4th International Workshop on Artificial Intelligence and Cognition, New York City, New York.
[15] Golob, E.J., Venable, K.B., Scheuerman, J., Anderson, M.T., 2017. Computational modeling of auditory spatial attention, in: Cog Sci 2017, London, UK.
[16] Greenwood, P.M., Parasuraman, R., 1999. Scale of attentional focus in visual search. Perception & Psychophysics 61.
[17] Itti, L., Koch, C., 2001. Computational modelling of visual attention. Nature Reviews. Neuroscience 2, 194–203. doi:10.1038/35058500.
[18] Lavie, N., 2005. Distracted and confused?: selective attention under load. Trends in Cognitive Sciences 9, 75–82. doi:10.1016/j.tics.2004.12.004.
[19] Lewandowsky, S., Farrell, S., 2010. Computational Modeling in Cognition: Principles and Practice. SAGE Publications.
[20] Mackworth, 1948. The breakdown of vigilance during prolonged visual search. The Quarterly Journal of Experimental Psychology 1, 6–21. doi:10.1080/17470214808416738.
[21] Mangun, G.R., Hillyard, S.A., 1988. Spatial gradients of visual attention: behavioral and electrophysiological evidence. Electroencephalography and Clinical Neurophysiology 70, 417–428.
[22] Muller, N.G., Mollenhauer, M., Rsler, A., Kleinschmidt, A., 2005. The attentional field has a Mexican hat distribution. Vision Research 45, 1129–1137. URL: http://www.sciencedirect.com/science/article/pii/S0042698904005735, doi:10.1016/j.visres.2004.11.003.
[23] Pillsbury, W.B., 1908. Attention. Half-title: Library of philosophy. Ed. by J. H. Muirhead, S. Sonnenschein & Co., ltd. The Macmillan co., London, New York. URL: http://catalog.hathitrust.org/Record/000384165.
[24] Rhodes, G., 1987. Auditory attention and the representation of spatial information. Perception & Psychophysics 42, 1–14. URL: http://link.springer.com/article/10.3758/BF03211508, doi:10.3758/BF03211508.
[25] Rorden, C., Driver, J., 2001. Spatial deployment of attention within and across hemifields in an auditory task. Experimental Brain Research 137, 487–496.
[26] Rossi, F., van Beek, P., Walsh, T. (Eds.), 2006. Handbook of Constraint Programming. Elsevier Science Inc., New York, NY, USA.
[27] Spence, C., Driver, J., 1997. Audiovisual links in exogenous covert spatial orienting. Perception & Psychophysics 59, 1–22. URL: http://link.springer.com/article/10.3758/BF03206843, doi:10.3758/BF03206843.
[28] Vera, A., Howes, A., McCurdy, M., Lewis, R., 2004. A constraint satisfaction approach to predicting skilled interactive cognition., in: Proceedings of the 2004 Conference on Human Factors in Computing Systems (CHI 2004), pp. 121–128.
[29] Wachtel, P.L., 1967. Conceptions of broad and narrow attention. Psychological Bulletin 68, 417–429.
[30] Zatorre, R.J., Mondor, T.A., Evans, A.C., 1999. Auditory attention to space and frequency activates similar cerebral systems. NeuroImage 10, 544–554. doi:10.1006/nimg.1999.0491.