A Reinforcement Learning Model Explains the Stagewise Development of Gaze Following

Hector Jasso[1], Jochen Triesch[2], Christof Teuscher[2]

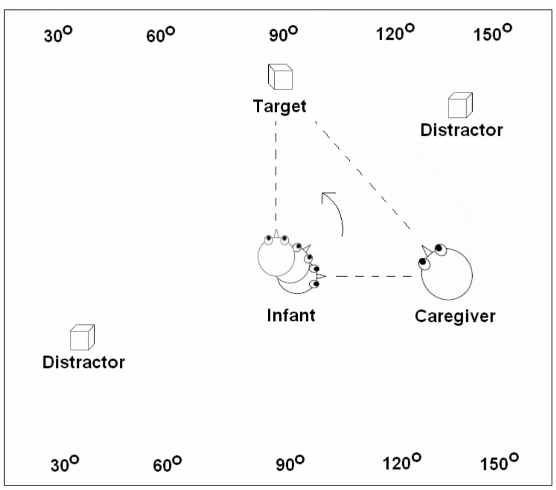
[1] Department of Computer Science and Engineering, University of California, San Diego {hjasso@cs.ucsd.edu}
[2] Department of Cognitive Science, University of California, San Diego {triesch@cogsci.ucsd.edu, christof@teuscher.ch}

Introduction

Gaze following is the ability to redirect one's visual attention towards an object that someone else is looking at. It is considered the cornerstone of social communication, for it allows infant individuals to integrate with their conspecifics by sharing references. We present a model of the development of gaze following based on reinforcement learning [1], where the infant learns to decode the caregiver's gaze by maximizing visual reward. The model emulates the stages that human infants go through, where gaze following is first learned for objects positioned in front of the infant, and only later for objects behind the infant. The model also learns to overcome the "Butterworth error" [2] [3], present in younger infants, where gaze is followed to the correct side of the room but sometimes stops at distractor objects positioned along the visual path between the caregiver and the target. This work was done as part of the MESA (Modeling the Emergence of Shared Attention) project at the University of California, San Diego [4] [5].

Room Setup and Reinforcement Learning Model

The infant is positioned at the center of a 42x27 unit room facing the caregiver (see Figure 1, left). Visually salient target and distractor objects can be placed anywhere in the room. The infant faces the caregiver at the beginning of each trial, and the caregiver looks at the target throughout the trial.



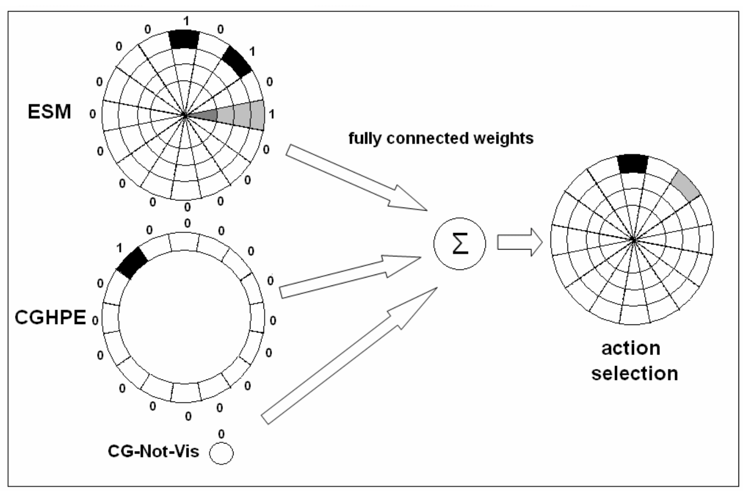


Figure 1. Left: Room setup. Infant following gaze in the presence of distractors. Right: Details of the model. Features calculated from the Expected Saliency Map (ESM), Caregiver Head Pose Estimate (CGHPE) and Caregiver Not Visible (CG-Not-Vis) inputs are weighted and added for each possible action. The action is selected using a softmax function. The activations shown correspond to the room configuration on the left figure when the infant is looking at the caregiver.

The heart of the model is a reinforcement learning algorithm (linear, gradient-descent Sarsa(λ)) (see Figure 1, right). **Input features** indicate the presence of objects (including caregiver) at 16 different headings from the infant's point of view, the estimate of the caregiver's head pose at 16 possible headings, and a feature indicating whether the caregiver is within the infant's field of view. 64 possible **actions** select the location (16 different headings, 4 different depths) where the infant can direct its attention. **Reward** is obtained by directing gaze to the location (both heading and depth) of an object. A fraction (40%) of this reward is obtained if gaze is directed to the wrong depth but to the correct heading of an object. Variable rewards can be specified for the caregiver, target object, and distractor objects. Additionally, a cost of 0 to 8 units is incurred at each time step for different gaze headings: 0 for looking straight ahead, 1 for looking at a heading of 22.5° to the left or right, 2 for looking at a heading 45° to the left of right, and so on.

Experiments and Results

In training trials, target and distractor objects are positioned randomly in the room, within a perimeter of r = 14 units from the infant. The infant is learns throughout 20 time steps, after which the learning trial is repeated. Once a specified number of training trials is completed, the infant is tested using a series of test trials. Successful test trials are defined as those where the infant shifts attention to the heading corresponding to the target during either one of time steps 10, 11, or 12. Three test setups are explored, where the last two are replicas of observational studies by Butterworth [3]: **Random object locations:** Target and distractor objects are positioned randomly in the room, within a perimeter of r = 14 units from the infant. **Target plus one distractor**: Target along the wall at 30^o, 60^o, 90^o, 120^o, and 150^o (see Figure 1, left), and a distractor object on the opposite side of the room. **Target plus 3 distractors ("testing for Butterworth error")**: Target along the wall at 30^o, 60^o, and 90^o, *first in the line of vision*, with a distractor in the same side of the room, at 90^o, 120^o and 150^o, respectively, and two distractors symmetrically positioned at the other side of the room. Target along the wall at 90^o, 120^o, and 150^o, *second in the line of vision*, with a distractor in the same side of the room, at 30^o, 60^o, and 90^o, respectively, and two distractors positioned symmetrically at the other side of the room.

The model was trained with saliency values for target and distractors of 50, saliency value for caregiver of 5. The learning rate α was set to 0.0005, the discount rate γ was set to 0.2, and the softmax temperature τ was set to 0.2. During each epoch, the model was trained for 500 trials. All experiments were successfully learned (had over 90% success rate at the end of epoch 50). The progression from learning to follow gaze to objects in front of the infant to those at the back matched that found in experimental observations by Butterworth [3]. Figure 2 left, shows the percentage of trial successes for different target locations for early and late epochs. Figure 2 right, shows details of how the Butterworth error is overcome.

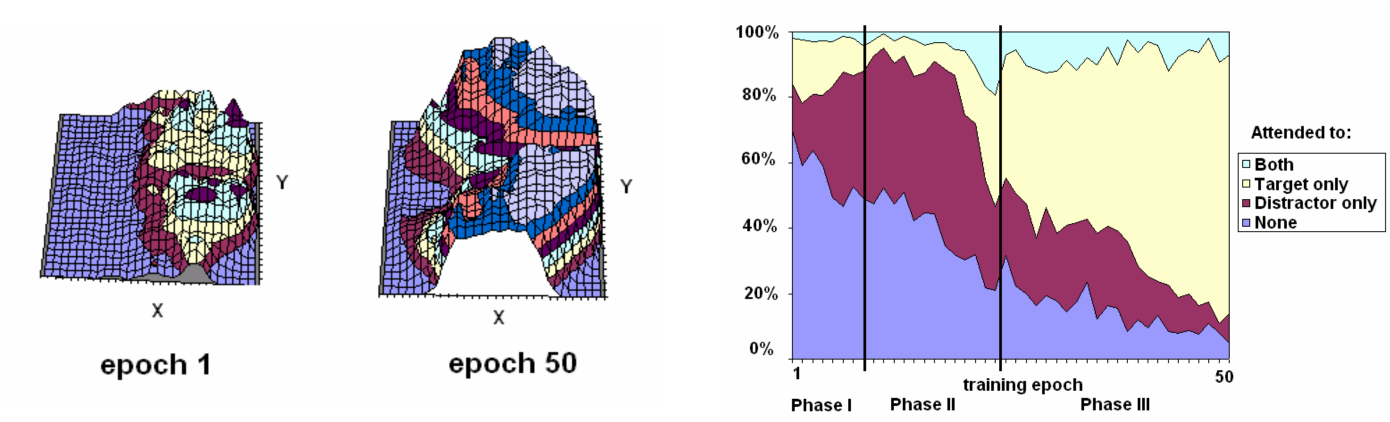


Figure 2. Left: Spatial distribution of trial successes. The grid specifies the location of the target. Infant's location corresponds to the center of the grid, facing right. References to objects in front of the infant are learned first (epoch 1), and those to objects behind are learned later (epoch 50). Right: Overcoming the Butterworth error. During Phase I of learning, neither the target nor the distractor was attended. In Phase II, the distractor is attended to. During Phase III, the infant overcomes the Butterworth error by looking at the target and ignoring the distractor.

Conclusions and Further Work

The model presented gives a parsimonious account of gaze following in a spatial setting. It is based on reinforcement learning, which has been proposed by the neuroscience community as a model of learning in the brain. Results obtained match the progression found in infants, where gaze following to objects in front of the infant are learned first and only later for objects at the back. They also match an observed early difficulty in overcoming the Butterworth error, and its solution at later stages of development. The model is currently being adapted to work in a more realistic environment using a virtual reality platform [6]. This will enable more sophisticated experiments involving three dimensions and realistic visual inputs.

References

[1] Sutton, R.S., and Barto, A.G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.
[2] Lau, B., and Triesch, J. *Learning Gaze Following in Space: a Computational Model.* 3rd International Conference for Development and Learning, ICDL'04, La Jolla, California, USA, October 20-22, 2004.
[3] Butterworth, G., & Jarrett, N. (1991). *What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy*. British Journal of Developmental Psychology, 9, 55-72.
[4] MESA project, http://mesa.ucsd.edu
[5] Fasel, I., Deak, G. O., Triesch, J. and Movellan, J. R. *Combining embodied models and empirical research for understanding the development of shared attention*. 2nd International Conference on Development and Learning, ICDL'02, Cambridge, MA, USA, June 12-15, 2002.
[6] Jasso, H., and Triesch, J. *A Virtual Reality Platform for Studying Cognitive Development*. 3rd International Conference for Development and Learning, ICDL'04, La Jolla, California, USA, October 20-22, 2004.