Towards a model of expectation-driven perception

Human perception is an active process by which meaningful information is gathered from the external environment. Application areas such as human-computer interaction (HCI), or the role of human experts in image analysis, highlight the need to understand how humans, especially experts, use prior information when interpreting what they see. Here, we describe how a model of expert perception is currently being extended to support expectation-driven perception of bitmap-level image data, focusing particularly on its ability to learn semantic interpretations.

The chrest model

CHREST (Chunk Hierarchy and REtrieval STructures[1]) is a computational model of perception and learning, designed to capture the perceptual knowledge acquired by an expert[2] (see also Gobet's individual article on page 8). Figure 1 illustrates the model's three main components: mechanisms for interacting with the external world; multiple short-term memories (STMs) to hold information from different input modalities; and a long-term memory (LTM) where information is held within a discrimination structure known as a 'chunking network'.

Recent work with CHREST is attempting to integrate three key processes for using expectations in perception: the use of bitmap data (whereas previous work has relied upon symbolic information), the creation of links between visual and verbal information, and the role of heuristics to guide the simulated eye. We describe the latter two in more detail here.

Combining visual and verbal chunks

CHREST's LTM holds information in the form of 'chunks', each of which is a familiar pattern in the environment. CHREST stores a chunk in a dual fashion. Firstly, the chunk itself is stored in a format representative of the data within it: in a visual domain, the chunk may be in the form of a bitmap; for a verbal pattern, it may be a sequence of phonemes. Secondly, the chunk's location in the model's LTM may be addressed directly with a link. Links are formed between nodes in the multiple STMs when they share an important relationship: such as being present in the environment simultaneously.

Figure 2 illustrates how a chunk acquired visually may be named by forming an association with a chunk acquired verbally. There are three steps. First, the visual pattern is sorted through LTM, and a pointer to the node retrieved is placed into visual STM. Second, the verbal pattern is sorted through LTM, and a pointer to the node retrieved is placed into verbal STM. Finally, a 'naming link' is formed between the two nodes at the top of the STMs.

Simulations with the CHREST model using semantic associations, such as those illustrated in Figure 2, demonstrate that CHREST captures several important phenomena illustrating the role of expectations in perception. These include: improved classification accuracy, faster classification, and the use of reconstructive memory to identify very noisy objects.[3]

Heuristics to guide eye fixations

An extended bitmap image cannot be perceived in its entirety. Instead, CHREST uses a simulated eye directed at a focus of attention—the fixation point—and has a limited field of view. The position of the eye is controlled with a set of heuristics that interact with each other. Prior work[2] has used various groups of heuristics that combine both bottom-up and top-down sources of information to guide the eye. In the top-down category, CHREST attempts to complete information held at a node referenced in the STM, to follow a test link, or to deepen the search within the LTM. Additional sources of information/heuristics include salient objects, novel objects, or default scanning of the scene.

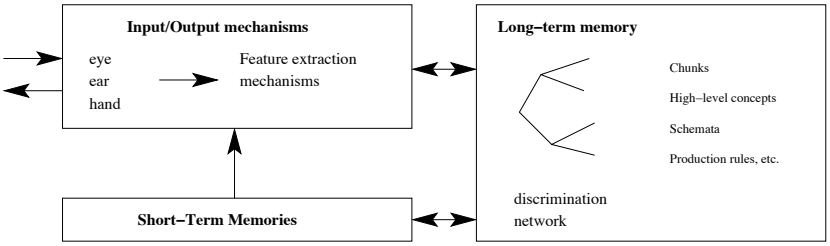

Figure 1. The CHREST Model.

CHREST is uniquely placed as a cognitive model of human learning in perceptual domains, with each area of Figure 1 interacting closely to gather and use meaningful information from a complex environment. With its recent extensions and use in domains with bitmap-level data, CHREST is currently being applied to domains involving the semantic analysis of complex images.

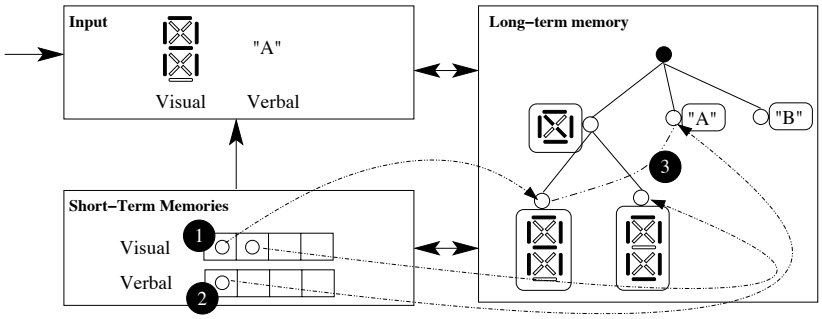

Figure 2. Learning a 'naming link' across two modalities.

Peter Lane* **and Fernand Gobet****
*Dept. of Computer Science
University of Hertfordshire, UK
E-mail: peter.lane@bcs.org.uk
http://homepages.feis.herts.ac.uk/~comqpcl
**Dept. of Human Sciences, Brunel University, UK
Details on page 8.

References
1. F. Gobet, P.C.R. Lane, S.J. Croker, P.C-H. Cheng, G. Jones, I. Oliver, and J.M. Pine, *Chunking Mechanisms in Human Learning,* **Trends in Cognitive Science 5** (6), pp. 236-43, 2001.
2. A.D. de Groot, F. Gobet, **Perception and Memory in Chess: Heuristics of the Professional Eye,** Van Gorcum, Assen, 1996.
3. P.C.R. Lane, A.K. Sykes, and F. Gobet, *Combining Low-Level Perception with Expectations in CHREST,* **Proc. of the European Cognitive Science Conf.,** 2003.