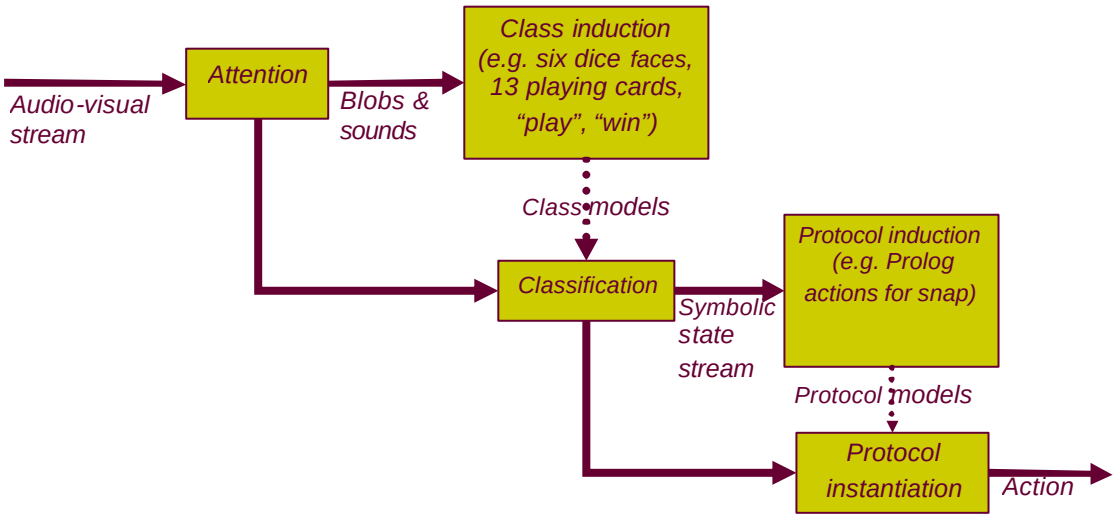
## *Deliverable 3.6*

# *Framework for the integration of statistical and logic-based models of objects and events*

The aim of Task 3.4 in Workpackage 3 has been to explore the integration of quantitative and qualitative spatio-temporal models. The framework presented in this deliverable is an outcome from this work. In developing the framework we have been guided by the goal of building a system that can learn to play simple table top games involving unfamiliar objects. This has proved to be a challenging task, involving 'end-to-end' learning of perceptual categories and game protocols – integrating qualitative and quantitative models in a single system. The deliverable is in seven parts: this introduction, four reports (Report1-4), and two packages of code.

The overall framework is described in Report 1 [1] (see also an earlier version in [5]) and summarised here with reference to the three other reports included with this deliverable.

The framework (illustrated below) is based on two learning phases in which the perceptual categories and game protocols are learnt, followed by an execution phase in which one of the players is simulated. These three phases are based around five distinct processes: attention, class induction, categorisation, protocol induction, and protocol instantiation.




In the first phase of learning, sounds and blobs are segmented from an audio-visual stream depicting repeated instances of the target activity (e.g. a card or dice game). Perceptual categories are learnt from these sounds and blobs, with the number of categories being chosen automatically, but with pre-determined types of category (e.g. based on colour, texture, and position). The details of this process are given in Report 2 [2] (see also [6]). Code for the perceptual learning is also included as part of the

deliverable. The segmentation and tracking of blobs was part of an earlier deliverable (see [7]).

In the second phase of learning, sounds and blobs are again generated from training data, and classified according to the categories learnt in the first learning phase. From the resulting stream of labelled sounds and blobs, game protocols are learnt using inductive logic programming. The details of this process are given in Report 3 [3], which will also appear at ECAI'04. This report demonstrates that the well-known problem of learning too many perceptual categories from audio-visual data is potentially solved by the formation of equivalence classes of categories that appear to be indistinguishable from the perspective of the game – they are used in the same way. Code for driving the inductive learning and interfacing to Progol is included as part of the deliverable.

The induction of game protocols in our demonstrator system is based on inverse entailment, embodied within the ILP system Progol. A comparative evaluation of Progol and the concept induction system HR is contained in Report 4 [4].

In the operational phase, the labelled sounds and blobs acquired from an audio-visual input stream, together with the learnt game protocol are used to predict the appropriate response for a player in each situation, and thereby substitute for that player in the game. We have developed a novel audio-visual rendering technique to simulate the utterances and facial expression of this player throughout the game.

In collaboration with The University of Genoa, we have begun to investigate the integration of this framework with a steerable camera in order to explore a layered attention mechanism integrating bottom-up saliency with top-down shifts of attention associated with a learnt task involving toy bricks. Details of this work have been included in Deliverable 2.4.

[1] Needham, C., D.R. Magee, V. Devin, P. Santos, A.G. Cohn and D.C. Hogg. **Autonomous learning of perceptual categories and symbolic protocols from audio-visual input**, submitted to BMVC'04.

[2] Magee, D., A.G. Cohn, D.C. Hogg. **Unsupervised Clustering, Feature Selection and Attention for Autonomous Visual Learning**, submitted to BMVC2004.

[3] Santos, P., D.R. Magee, A. Cohn and D. Hogg (2004, to appear). **Combining Multiple Answers for Learning Mathematical Structures from Visual Observations**. European Conference on Artificial Intelligence (ECAI'04), Valencia.

[4] Colton. S, P. Santos and D.R. Magee. **Predictive and Descriptive Approaches to Learning Game Rules from Vision Data**, submitted to ILP'04.

[5] Magee D., Needham C.J., Santos P., Cohn A.G. and Hogg D.C. (2004, to appear). **Autonomous learning for a cognitive agent using continuous models and inductive**

**logic programming from audio-visual input**, Proc. AAAI Workshop on Anchoring Symbols to Sensor Data.

[6] Magee D. (2003) **A Sequential Scheduling Approach to Combining Multiple Object Classifiers Using Cross-entropy**. Proc. IAPR International Workshop on Multiple Classifier Systems, (Springer LNCS 2709), 135-145.

[7] Magee D. (2004). **Tracking Multiple Vehicles using Foreground, Background and Motion Models**. Image and Vision Computing, 22(2), 143-155.