Specifying Modalities in the MGLAIR Architecture

Jonathan P. Bona and Stuart C. Shapiro

Department of Computer Science and Engineering
University at Buffalo, The State University of New York
{jpbona,shapiro}@buffalo.edu

Abstract. The MGLAIR cognitive agent architecture includes a general model of modality and support for concurrent multimodal perception and action. An MGLAIR agent has as part of its implementation multiple modalities, each defined by a set of properties that govern its use and its integration with reasoning and acting. This paper presents MGLAIR's model of modality and key mechanisms supporting their use as parts of computational cognitive agents.

Keywords: multi-modality, cognitive architectures, agent architectures, embodied agents

1 Introduction

The MGLAIR (Multimodal Grounded Layered Architecture with Integrated Reasoning) cognitive agent architecture extends the GLAIR architecture [1] to include a model of concurrent multimodal perception and action. Of central importance to MGLAIR is its treatment of afferent and efferent *modalities* as instantiable objects that are part of agent implementations. The architecture specifies how modalities are defined and managed, what properties they posses, and how their use is integrated with the rest of the system. Agents using these modalities deal independently with sense data and acts that correspond to distinct capabilities.

MGLAIR is divided into three major layers, illustrated in Figure 1. The Knowledge Layer (KL) and its subsystems perform conscious reasoning, planning, and acting. A gradation of abstractions across the layers of the architecture terminates in symbolic knowledge at the KL. An MGLAIR agent becomes consciously aware of percepts when they are added to its KL. The Sensori-Actuator Layer (SAL) is embodiment-specific and includes low-level controls for the agent's sensori-motor capabilities. The Perceptuo-Motor Layer (PML) connects the mind (KL) to the body (SAL), grounding conscious symbolic representations through perceptual structures. The PML is further stratified into sub-layers. The highest PML sub-layer, comprised of PMLa and PMLs, grounds KL symbols for actions and percepts in subconscious actions and perceptual structures respectively. The lowest PML sub-layer, the PMLc, directly abstracts the sensors and effectors at the SAL into the basic behavioral repertoire of the robot body. The middle PML layer, the PMLb, handles translation and communication between the PMLa and the PMLc. The inward-pointing and outward-pointing arrows spanning the layers in Figure 1 represent afferent and afferent modalities respectively.

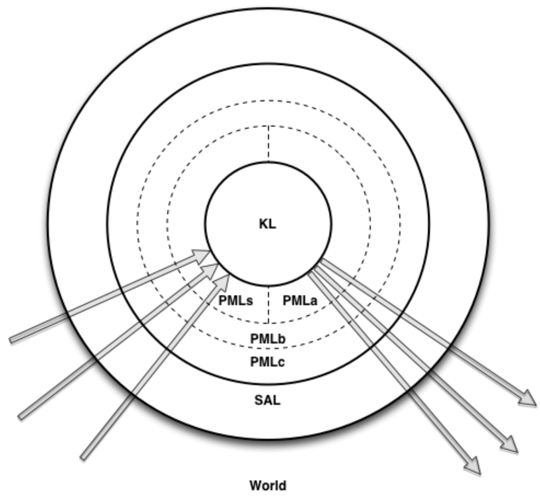

Fig. 1. MGLAIR

MGLAIR's KL is implemented in SNePS, a logic-based Knowledge Representation and Reasoning system [2] [3]. SNeRE (the SNePS Rational Engine), the SNePS subsystem that handles planning and acting [4], is a key component of MGLAIR. SNeRE connects agents' logic-based reasoning with acting. The plans formed and actions taken by an agent at any time depend in part on the agent's beliefs (including beliefs about the world based on its perceptions) at that time.

2 Modality

A modality corresponds to a single afferent or efferent capability of an agent: a limited resource capable of implementing only a limited number of related activities simultaneously. Each agent's embodiment determines which modalities are available to it.

An MGLAIR modality possesses a directional data channel that connects the mind with the body. The modality itself handles the transmission and transformation of data in the channel. Within an afferent modality raw sensory data originating at the SAL is passed up to the PML and converted to perceptual structures that are aligned with and used as the basis for conscious symbolic percepts at the KL. Any action an agent consciously performs is available as an abstract symbolic representation in the knowledge layer, corresponding to low-level motor control commands in the SAL, which it is connected to through alignment with the intermediate PML structures. The flow of distinct types of sensory and motor impulses between the agent's mind and body occur independently, each in its own modality. MGLAIR's PML structures constitute unconscious multi-modal representations. Though they are inaccessible to the agent for conscious reflection, they play a crucial role in its cognition.

2.1 Modality Properties

An MGLAIR modality specification is a 9-tuple of modality properties, which determine its behavior:

$\langle$ `name, type, pred, chan, access, focus, conflict, desc, rels` $\rangle$:

These are: a unique *name* for the modality; a *type*, (subtype of `afferent` or `efferent`); KL *predicates* used for percepts or acts in the modality; the modality's *data channel*; a flag granting/denying the agent conscious *access* to the modality; a specification for modality *focus* (see §3.5); a *conflict* handler for when multiple acts or multiple sensations try to use the modality simultaneously (see §3.2, §3.4); a *description*; and *relations* to other modalities.

3 Key Modality Structures and Mechanisms

3.1 Perceptual Functions

Each afferent modality has a function specific to it that is applied to perceptual structures in its PMLs to convert them into symbolic knowledge and assert them to the KL. The nature of the perceptual function depends on the type of perceptual structures it handles, and on the nature of the sensor and modality. For instance, a perceptual function for a visual modality might take as input a structure representing different visual features (shapes, colors, etc), and produce KL terms and propositions classifying and identifying objects in the field of vision. All perceptual functions take as input a timestamped PML structure from the modality's perceptual buffer and produce as output KL terms that represent the percept using predicates associated with the modality. Each modality's perceptual function also links the resulting terms to the modality in which they originated by embedding a reference to the modality within the SNePS data structure that represents each term.

3.2 Perceptual Buffers

Modality buffers for each perceptual modality in the PMLs queue up perceptual structures to be processed and consciously perceived by the agent. Perceptual buffers may have a fixed capacity that limits the number of elements the buffer can hold. Otherwise, buffers with unlimited capacity must have an expiration interval, an amount of time after which data in the buffer expires and is discarded without being perceived by the agent.

Modality buffers serve to smooth out perception and minimize the loss of sensory data that the agent would otherwise fail to perceive because of temporary disparities between the speed with which the sensor generates data and the time it takes to process and perceive that data. By adjusting a buffer's size or expiration interval, and by specifying how full buffers are handled, an agent designer may achieve a range of possible effects suitable for different types of agents and modalities.

Timestamped sensory structures assembled at the PMLb are added to the modality buffer as they are created. If the buffer is full when a structure is ready to be added, for instance because the buffer has a fixed capacity and the expiration interval is too long, then an attempt to add a new structure to the buffer will be handled either by blocking and discarding the new data rather than adding it to the buffer - in which case it will never be perceived by the agent - or by making space for the new data by deleting, unperceived, the oldest unprocessed structure in the buffer even if it has not expired.

Each modality buffer has a buffer management process that repeatedly removes the oldest non-expired structure from the buffer and applies the modality's main perceptual function to it. These processes are affected by changes in the modality's focus, discussed more in §3.5.

3.3 Act impulses in efferent modalities

At the knowledge layer, SNeRE connects the agent's reasoning and acting capabilities through the management of policies and plans. An example policy (stated in English from the agent's perspective) is, "Whenever there is an obstacle close in front of me, I should move back, then turn, then resume operating." An example plan is, "To pick up an object, I first open my grasping effector, then position it over the object, then I lower it, then I grasp with it, then I raise it." Each act plan consists of a *complex act*, which is being defined by the act plan, and a sequence of other acts that comprise the complex act. These may be themselves complex acts or primitive acts. All such plans must eventually bottom out in *primitive acts* – acts that the agent cannot introspect on or further divide into their composite parts, but which it may simply `perform`.

Each efferent modality contains within its PML an action buffer that stores act impulses resulting from conscious actions the agent has performed, but that have not yet been executed at the lower levels (i.e. they have not yet caused the relevant effectors to have their effects on the world).

Plans for complex acts may be comprised of acts that use different modalities. For instance, depending on the agent's embodiment, moving a grasper and using it may be independent of each other, and therefore use separate modalities. Since these separate modalities operate independently, actions may occur in them simultaneously. For instance, *moving the grasper arm to a position* and *opening the grasper*, if they use separate motors, could very well be performed at the same time without effecting each other's operation.

3.4 Efferent Modality Buffers

Like percept buffers, action buffers are located in the PMLb. Act impulses are added to the buffer as a result of primitive acts that are performed at the PMLa, and are removed and processed at the PMLc, where they are further decomposed into low-level commands suitable for use by the SAL. For instance, a primitive action for a grasping effector might be to move the effector some number of units in a particular direction. When the PML function attached to this action is called, it places a structure representing this impulse and its parameters into the modality's action buffer. As soon as the modality is available (immediately, if it was not already carrying out some action when the `move` action was performed), the structure is removed from the buffer and processed at the PMLc, where it is converted into commands the SAL can execute (e.g. apply a certain amount of voltage to a particular motor for a duration).

When an efferent modality's act buffer is empty, an act impulse added to the buffer as a result of the agent's consciously performing an act will be immediately removed and sent to the SAL for execution. When an action is performed using an efferent modality whose buffer is not empty, the default behavior is to add the act impulse to the buffer, from which it is be removed and executed when the modality is available. For example, a speech modality might be configured to buffer parts of utterances and realize them in order as the relevant resource becomes available. Another option is for the new impulse to clear the buffer of any existing act impulses. For instance, an agent with a locomotive modality might be configured to discard outstanding impulses and navigate to the location it has most recently selected. The manner in which such conflicts are resolved is specified as part of the modality specification.

Action buffers in efferent modalities have similar properties to the percept buffers in afferent modalities: their capacities are configurable as are their expiration intervals. One difference is that MGLAIR's model of modality focus currently applies only to perception and does not account for focusing on actions in efferent modalities.

3.5 Regulating Modality Focus

MGLAIR's model of modality focus allows agents to selectively dedicate more or less of their resources to the task of perceiving within a particular afferent modality. An agent designer may specify a default focus level for a modality and either permit or forbid the agent to adjust its focus levels on its own modalities – including the possibility of ignoring a modality altogether. Altering a modality's focus increases or decreases the frequency with which its internal processes run. This allows an agent to prioritize processing of percepts from a particular modality relative to the others depending on its current task and priorities.

Because each agent has only a limited amount of computational resources it is possible for a modality to interfere with the operation of others by monopolizing those resources, even though each modality is operating independently with its own processes and structures. This happens when a modality has to work harder than others to convert its sensory data into perceptual knowledge, or when one modality is producing much more, or much richer, sensory data than others. Without a mechanism to selectively allocate its focus to the modality that is more relevant to its situation and task, agents would expend most of their processing power on more demanding modalities, even when they are less urgent or less relevant to the agent's task than some less demanding modalities.

Altering a modality's focus adjusts the priority of the modality's buffer management and perceptual function processes. For an agent with just two modalities with nearly all of the same properties connected to two identical sensors in the same environment, but with different levels of focus, the modality with the lower focus will remove and process PML structures from its buffer less frequently, and will therefore be more likely to reach its capacity or to have elements reach their expiration times before they are processed and perceived. That is, an agent that is not focused on some modality may fail to perceive some things that the modality is capable of sensing.

4 Example Modality

As an example, consider an agent embodied in a small robot with - among other sensors - a unidirectional ultrasonic range finder that the agent uses to avoid bumping into obstacles as it moves around its environment. The agent may be equipped with other sensors and effectors as well – a camera based visual modality, motorized wheels, a grasping effector.

To make use of this sensor, the agent must have a way of representing the information that the sensor conveys and of becoming aware of (perceiving) it. It becomes aware of percepts via an afferent modality that connects the sensor to the agent's mind, with mechanisms to convert raw sensory data into the agent's perceptual representation. The range finder operates by sending out sound waves and then detecting how long it takes for the echo to bounce off of the nearest object and return to the sensor. At the lowest level - the SAL - the data produced by this sensor is in the form of voltage levels generated by a piezoelectric receiver. These values are converted into an integer value between 0 and 255, which corresponds to the distance to the nearest object in front of the sensor. To use this sense to avoid banging into walls and other obstacles, it may suffice for the agent to perceive that the next thing in front of it is very close, very far, or in-between, rather than a precise value. This is achieved by having the PML convert values like `34, 148`, etc, into structures at the granularity we want the agent to perceive: `close`, `medium`, and so on. These are removed from the modality's sensory buffer and processed by the modality's perceptual function, which produces symbolic representations that are then asserted at the KL, e.g. the term `DistanceIs(far)`, which represents the proposition that the distance to the nearest object in front of the agent is far. These percepts (beliefs) affect the agents behavior if it holds policies like "when the nearest thing in front of me is very close, stop any forward motion and turn before resuming it." Figure 2 shows the layers and examples of information present at each.

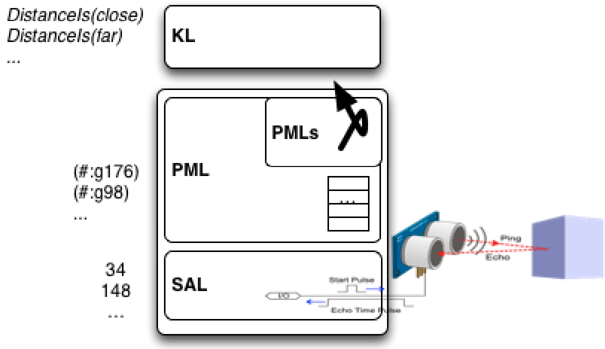

Fig. 2. View of a single afferent modality for an ultrasonic sensor [1]

5 Conclusions

The MGLAIR architecture provides a model of afferent and efferent modalities for computational cognitive agents. MGLAIR agents that instantiate these modality objects can use them to sense and act - and make inferences and plans about percepts and actions - concurrently in different modalities.

By dividing agents' capabilities into modular modalities, each with its own properties governing its use, MGLAIR allows agents to sense and act simultaneously using different resources with minimal interference, and to consciously decide which resources to focus on for particular tasks. The specific properties that determine how each modality functions, are defined as part of a modality specification that is shared among the layers of the architecture.

References

1. Shapiro, S.C., Bona, J.P.: The GLAIR Cognitive Architecture. International Journal of Machine Consciousness **2**(2) (2010) 307–332
2. Shapiro, S.C., Rapaport, W.J.: The SNePS family. Computers & Mathematics with Applications **23**(2–5) (1992) 243 – 275
3. Shapiro, S.C., The SNePS Implementation Group: SNePS 2.7 User's Manual. Department of Computer Science and Engineering, University at Buffalo, The State University of New York, Buffalo, NY. (2007) Available as `http://www.cse.buffalo.edu/sneps/Manuals/manual27.pdf`.
4. Kumar, D.: A unified model of acting and inference. In Nunamaker, Jr., J.F., Sprague, Jr., R.H., eds.: Proceedings of the Twenty-Sixth Hawaii International Conference on System Sciences. Volume 3. IEEE Computer Society Press, Los Alamitos, CA (1993) 483–492

[1] Sensor diagram ©Parallax Inc., available under the Creative Commons Attribution-Share Alike 3.0 license at `http://learn.parallax.com/KickStart/28015`