Modality in the MGLAIR Architecture

Jonathan P. Bona* and Stuart C. Shapiro †

Abstract The MGLAIR cognitive agent architecture includes a general model of modality and support for concurrent multimodal perception and action. It provides afferent and efferent modalities as instantiable objects used in agent implementations. Each modality is defined by a set of properties that govern its use and its integration with reasoning and acting. This paper presents the MGLAIR model of modalities and mechanisms for their use in computational cognitive agents.

1 Introduction

The MGLAIR (Multimodal Grounded Layered Architecture with Integrated Reasoning) cognitive agent architecture extends the GLAIR architecture [10] to include a model of concurrent multimodal perception and action.

Humans constantly sense and act in multiple modalities simultaneously. MGLAIR aims to replicate this functionality in computational agents by including modular afferent and efferent modalities as part of the architecture – without necessarily mimicking the actual implementation of the human mind.

As a cognitive architecture [7], MGLAIR provides a high-level specification of a system in terms of its parts, their functionality, and the interactions between parts. This system may be used as a platform to implement and test models of cognition.

MGLAIR is a layered architecture: the Knowledge Layer (KL) and its subsystems perform conscious reasoning, planning, and acting. Each agent's Sensori-Actuator Layer (SAL) is embodiment-specific and includes low-level controls for its sensori-motor capabilities. The Perceptuo-Motor Layer (PML) connects the mind

———————————

Jonathan P. Bona
University at Buffalo Computer Science and Engineering, e-mail: jpbona@buffalo.edu

Stuart C. Shapiro
University at Buffalo Computer Science and Engineering, e-mail: shapiro@buffalo.edu

(KL) to the body (SAL), grounding conscious symbolic representations through perceptual structures. The PML is itself divided into sub-layers.
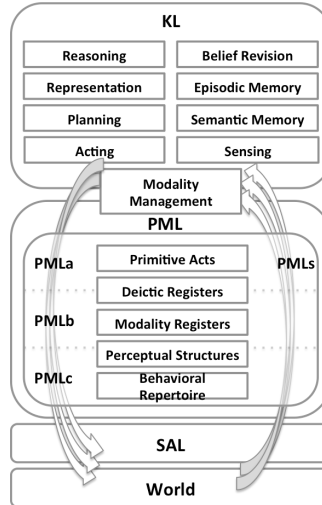

Fig. 1: MGLAIR

MGLAIR's Knowledge Layer is implemented in SNePS, which is a logic-based Knowledge Representation and Reasoning system [11] [12]. SNeRE, the SNePS subsystem that handles planning and acting [5], is a key component of MGLAIR. It expands agents' capabilities by connecting 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.

SNeRE includes mental acts (*believing* or *disbelieving* a proposition; *adopting* or *unadopting* a policy), control acts (control structures such as act *sequence* and *selection*), and primitive, agent-specific external acts such as moving, sensing, or other interactions with the world. Sensing and acting within MGLAIR involve the flow of data between the agent's conscious mind and its body. When, in the course of reasoning and interacting with the environment, an agent decides to move, it does so by performing actions that utilize the appropriate modality.

Though some existing GLAIR agents use multiple sensory and motor capabilities [8][9], and were initially considered to be "MGLAIR agents," their multiple modalities are implemented *ad hoc* in software and lack an underlying structure and support from, or integration with, the architecture or the agent as a whole.

Of central importance to MGLAIR is its treatment of afferent and efferent *modalities* as instantable 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.

2 Modality

Each 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.

The set of modalities available to an agent necessarily depends on its embodiment: an agent with motorized wheels and a camera-based vision system moves by consciously performing actions that use its locomotion modality, and consciously perceives the world through the use of its visual modality.

Other architectures (e.g. ACT-R [1], Soar [6], CERA [2]) include processing modules for sensory and motor capabilities/modalities. These are usually fixed in number and in nature, even in architectures such as EPIC [4], which focuses on support for multimodal embodied tasks. While MGLAIR provides default modality definitions for commonly-used modalities, its general model allows agent designers to extend, exclude, or modify these, and to define as many additional modalities as are needed for different agents and embodiments. It provides sensible, customizable, default procedures for the integration of percepts within the knowledge layer.

In MGLAIR, all of an agent's conscious thought takes place in the agent's symbolic language of thought at the KL. MGLAIR's PML structures, with which symbolic knowledge is aligned, constitute unconscious multi-modal representations.

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.

An MGLAIR modality possesses a directional data channel that connects the mind to the body. The modality itself handles the transmission and transformation of data in the channel: raw sense data within a modality at the SAL is passed up to the PML and converted to perceptual structures. These structures in the PML are converted to conscious symbolic representations of percepts as they are passed upwards to the KL. The exact nature and implementation of that transformation depends on the type of modality in question and on the agent's embodiment.

The SNePS KL and its connections to the PML are implemented in Common Lisp. MGLAIR does not require the PML itself or the SAL to be implemented using any particular technology: the software implementations of an agent's KL, PML, and SAL may even run on different physical machines.

2.1 Modality Properties

An MGLAIR modality definition is a set of **required** and *optional* properties:

A unique name, by which the agent may refer to the modality;
A modality type, a subtype of either `afferent` or `efferent`;
Knowledge layer predicates for symbolic representation of percepts or acts;

Data channel specification, including connection type and parameters, as well as a **modality buffer specification**, discussed more in Section 2.2;
Conscious access permissions granting the agent knowledge of the modality;
Focus permissions governing focus for the modality (see §2.2);
A conflict handling specification governing conflicting access attempts (§2.4);
A human-readable description of the modality;
Relations to other modalities – e.g. for proprioception, discussed in §2.5.

Each modality's definition is shared among the layers of the architecture.

2.2 Modality Buffers

Each afferent modality data channel includes a buffer of sense data arriving from the SAL via the lower levels of the PML. Sense data are added to the buffer as they are generated, and removed in the same order for processing by the upper levels of the PML. Sense data in the modality buffers are marked with timestamps. Neither the raw sense data nor their timestamps are consciously accessible to the agent, though both can be used within the PML to construct percepts.

To prevent agents from becoming overwhelmed by sense data generated faster than they can be perceived, modality buffers have an optional fixed capacity, which is specified when the modality is defined. A modality with a fixed capacity may handle incoming sense data that would otherwise be placed in its (full) buffer either by refusing and discarding arriving sense data when the buffer is full, or by pushing out old data to make room for the new. Modality buffers allow for either of these possibilities.

Modality buffers without a fixed capacity must specify an expiration interval for data. Old data in a modality buffer that reaches expiration are discarded without being processed/perceived. A buffer may have both a capacity and an expiration interval.

2.3 Modality Focus

An MGLAIR agent may elect to focus more on a particular afferent modality than on others, or to ignore a particular modality altogether. Each distinct modality has its own processes in the PML that operate continually and concurrently to manage the flow and transformation of data/impulses in the modality. Altering a modality's focus increases or decreases the frequency with which its internal processes run.

The ability to govern modality focus is useful when an agent is carrying out a task that relies heavily on perceiving with a subset of its sensory modalities, or when percepts in some modalities may be more urgent than percepts in others. An agent embodied in a robot with cliff detection and speech recognition might focus more on perceiving with its cliff detection modality than on its speech recognition.

In this case, and in many others, the less urgent modality may have a much higher volume of sense data than the more urgent one. Without a focusing mechanism, such an agent would dedicate its attention to the modality producing the most percepts.

An agent may alter its focus on a modality by performing one of the following acts with the modality name as its argument: `focus` increases the agent's focus on the modality; `unfocus` decreases focus; `ignore` causes the agent to ignore the modality completely; `attend` causes the agent to stop ignoring the modality, restoring it to the focus level it held before being ignored; and `restore-focus` resets the modality's focus to its starting level.

In order for an agent to consciously act to modify its focus on a modality, the agent must be aware of the modality, and must have a way of referring to it. Not all modalities must be consciously accessible to their agents. Whether a modality is accessible to an agent is a property of the modality, specified in its definition.

By default, all of an agent's modalities have equal focus levels. Each modality definition may include an initial level of focus for the modality. A modality definition may also prohibit the agent from changing a modality's level of focus.

2.4 Sensing and Acting

Rules in the upper layer of each afferent modality transform perceptual data from the PML into conscious knowledge within the KL. This involves converting perceptual structures into propositional terms for what is being perceived. These rules and structures are implemented in PMLs: the PML, sub-layer *s*. Each such rule is specific to its modality and uses predicates and terms that may be unique to the modality to generate the propositional terms that correspond to percepts in that modality.

Though it is up to the agent designer to specify how complex sense data are are translated into conscious percepts, the architecture includes default translations for each afferent modality. These defaults are activated by instantiating and using a modality that does not redefine them. Separate processes for each modality connect the PML and KL, converting sense data into knowledge concurrently as it arrives. This allows the agent to perceive via all of its afferent modalities simultaneously.

Percepts arriving at the KL can be linked to the modality that produced them, allowing the agent to reflect on how it came to believe that something is the case ("I saw it"), and to perform meta-reasoning about its own perceptual capabilities.

When a primitive act is defined it is associated with an efferent modality. A single modality may have associated with it multiple acts that use the same basic bodily capability or resource. For instance, an agent with a grasping effector might have a modality dedicated to its use. Distinct acts that use the modality, such as *positioning the grasper*, *grasping an object*, and *releasing an object*, have distinct conscious representations. When the agent performs acts, the PML and SAL decompose them from abstract symbols (e.g. `grasp`) into low-level commands (e.g. `apply voltage to motor m for n seconds`), and executes those commands.

Each efferent modality can carry out only a single activity at a time. Any desired activity that seems to violate this principle should be broken into component acts in separate modalities for concurrent use. Plans for complex acts are formed by combining other acts, either primitive or complex, via control acts

If an agent attempts to use a single efferent modality while it is already in use, the conflict is handled in one of several ways: The new impulse may be blocked and discarded; the new impulse may interrupt and discard the current activity, then proceed as normal; the new impulse may interrupt and suspend the current activity, then proceed as normal, in which case the interrupted activity resumes once the interrupter finishes; or the new impulse may be placed in an act queue for that modality and be carried out when the modality is next available. The manner in which such conflicts are handled is a property of each modality, and part of its definition.

2.5 Relations Between Modalities

We have implemented a simple model of proprioception for MGLAIR in which an efferent modality is associated with a related afferent modality, allowing the agent to perceive its own actions based on sense data generated in the SAL. For example, an agent that is embodied in a wheeled robot with motors that include a tachometer can use this feature to sense its use of its motors.

The modalities for *locomotion* and *locomotive proprioception* are connected at the PML, where a locomotive proprioception modality register stores perceptual data pertaining to the current state of locomotion. This may include a simple sense corresponding to its just being in use, or may be a more advanced sense based on the position of motors, the duration of the current use, etc. Depending on the application and design of the agent, the agent may be given conscious access to this register. In this case, the agent may simply know whenever its motors are in use because it is able to directly perceive that fact. Alternatively, proprioceptive modality registers may remain hidden in the PML with their contents inaccessible to conscious reasoning. In this case, the sense information they represent can still be used to manage act impulses for that modality within the PML, for instance by blocking new acts in the associated efferent modality while it is still in use at the SAL.

This example of an efferent modality tied to a corresponding afferent modality may serve as a model for the development of more complex inter-modal interactions. A similar approach might join senses from modalities before they are perceived. E.g., an agent with vision and sonar-based object detection might merge sensory data from those two modalities into joint percepts in the PML. This approach might be used to model many of the forms of cross-modal interactions observed in nature [13] [3].

3 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 resulting agents are more useful than counterparts that can be implemented without such modalities, and more accurately mimic the capabilities of cognitive agents observed in nature.

References

1. J. Anderson. ACT: A simple theory of complex cognition. *American Psychologist*, 51:355–365, 1996.
2. R. Arrabales, A. Ledezma, and A. Sanchis. A Cognitive Approach to Multimodal Attention. *Journal of Physical Agents*, 3(1):53, 2009.
3. G. Calvert, C. Spence, and B. Stein, editors. *The Handbook of Multisensory Processes*. MIT Press, 2004.
4. D. Kieras and D. Meyer. An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12(4):391–438, 1997.
5. D. Kumar. A unified model of acting and inference. In J. F. Nunamaker, Jr. and R. H. Sprague, Jr., editors, *Proceedings of the Twenty-Sixth Hawaii International Conference on System Sciences*, volume 3, pages 483–492. IEEE Computer Society Press, Los Alamitos, CA, 1993.
6. J. E. Laird, A. Newell, and P. S. Rosenbloom. Soar: an architecture for general intelligence. *Artif. Intell.*, 33(1):1–64, 1987.
7. P. Langley, J. Laird, and S. Rogers. Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, 10(2):141–160, 2009.
8. S. Shapiro, J. Anstey, D. Pape, T. Nayak, M. Kandefer, and O. Telhan. MGLAIR agents in virtual and other graphical environments. In *Proceedings of the National Conference on Artificial Intelligence*, page 1704. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.
9. S. C. Shapiro, J. Anstey, D. E. Pape, T. D. Nayak, M. Kandefer, and O. Telhan. The Trial The Trail, Act 3: A virtual reality drama using intelligent agents. In R. M. Young and J. Laird, editors, *Proceedings of the First Annual Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE-05)*, pages 157–158. AAAI Press, Menlo Park, CA, 2005.
10. S. C. Shapiro and J. P. Bona. The GLAIR Cognitive Architecture. *International Journal of Machine Consciousness*, 2(2):307–332, 2010.
11. S. C. Shapiro and W. J. Rapaport. The SNePS family. *Computers & Mathematics with Applications*, 23(2–5):243 – 275, 1992.
12. S. C. Shapiro and 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.
13. B. Stein and M. Meredith. *The Merging of the Senses*. The MIT Press, 1993.