Available online at www.sciencedirect.com
SCIENCE DIRECT®

Neural Networks 19 (2006) 272–284

Neural Networks

www.elsevier.com/locate/neunet

2006 Special issue

# Perceiving the unusual: Temporal properties of hierarchical motor representations for action perception

Yiannis Demiris *, Gavin Simmons

*Biologically Inspired Autonomous Robots Team (BioART), Intelligent Systems and Networks Group, Department of Electrical and Electronic Engineering, Imperial College London, South Kensington Campus, London SW7 2BT, UK*

**Abstract**

Recent computational approaches to action imitation have advocated the use of hierarchical representations in the perception and imitation of demonstrated actions. Hierarchical representations present several advantages, with the main one being their ability to process information at multiple levels of detail. However, the nature of the hierarchies in these approaches has remained relatively unsophisticated, and their relation with biological evidence has not been investigated in detail, in particular with respect to the timing of movements. Following recent neuroscience work on the modulation of the premotor mirror neuron activity during the observation of unpredictable grasping movements, we present here an implementation of our HAMMER architecture using the minimum variance model for implementing reaching and grasping movements that have biologically plausible trajectories. Subsequently, we evaluate the performance of our model in matching the temporal dynamics of the modulation of cortical excitability during the passive observation of normal and unpredictable movements of human demonstrators.

*Keywords:* Imitation; Minimum variance; Hierarchical structures; Action recognition; Corticospinal excitability

## 1. Introduction

An increased interest in computational mechanisms that will allow robots to observe, imitate and learn from human actions has resulted in a number of computational architectures that allow the matching of demonstrated actions to the observer robot's equivalent motor representations (Alissandrakis, Nehaniv, & Dautenhahn, 2002; Billard, 2000; Demiris & Hayes, 2002; Schaal, Ijspeert, & Billard, 2003). These architectures, whilst sharing common computational components such as modules for processing and classifying visual information and retrieving motor representations, differ in the way that the perceptual information is coded and classified, the organisation of the motor system, and the stage at which the motor representations are used. The final aspect, at what stage the motor representations are used, differentiate architectures that follow the general 'observe, classify, imitate' decomposition (Kuniyoshi, Inaba, & Inoue, 1994), from those that advocate a stronger involvement of the motor systems in the perception process, through a 'rehearse, predict, observe, reinforce' decomposition (Demiris & Hayes, 2002; Demiris & Johnson, 2003; Schaal et al., 2003). In the latter, the observer robot invokes its motor systems to rehearse potential actions, predicting and confirming incoming observed states during the demonstration. This approach has gained biological credibility with the discovery of the mirror system in monkeys and humans (Grezes, Armony, Rowe, & Passingham, 2003; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Not all theoretical models advocate the actual rehearsal of candidate actions as our previous work has done (Demiris & Hayes, 2002), opting instead for a weaker version of this motor theory of perception, usually termed 'motor resonance', in which the motor representations are retrieved through a resonance mechanism rather than a generative mechanism.

For imitation approaches that advocate the use of motor systems during the perception stage it becomes crucial to have a clear and flexible motor system organisation. Hierarchical representations, involving primitive motor structures at the lowest level, while increasing their complexity in higher levels, have been proposed (Demiris & Johnson, 2003; Wolpert, Doya, & Kawato, 2003), and tested in robotic systems (Demiris & Johnson, 2003), which successfully learned and used sequences of actions by observation. However, little has been done with respect to the temporal dimension of these representations, including how they can be coordinated, as well as their relation to biological data.

---

* Corresponding author.
*E-mail address:* y.demiris@imperial.ac.uk (Y. Demiris).
*URL:* http://www.iis.ee.ic.ac.uk/yiannis (Y. Demiris).

0893-6080/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2006.02.005

In this paper, we will examine in detail the issue of hierarchical representations, and in particular examine how higher level models can be composed from (and coordinate) lower levels primitives. Our approach will use representations based on the biologically plausible minimum variance model of movement control (Harris & Wolpert, 1998; Simmons & Demiris, 2005), which leads to principled and biologically plausible coordination of the underlying components. We subsequently compare a particular instantiation of our hierarchical attentive multiple models for execution and recognition (HAMMER) architecture (Demiris & Khadhouri, in press) for reaching and grasping actions, with transcranial magnetic stimulation (TMS) data from humans during the passive observation of grasping movements by a demonstrator (Gangitano et al., 2004).

## 2. Background

### 2.1. Hierarchies

Hierarchies are computationally interesting since they advocate a logical representational decomposition: motor primitives at the lower levels take care of the executional details while progressively higher levels shift their emphasis towards exerting temporal, contextual and cognitive control. From a robotics point of view, this allows for easier task planning and execution. In action understanding and gesture recognition, hierarchical representations have been regularly used since they allow the processing of information in multiple levels of detail (Hu, Tan, Wang, & Maybank, 2004). There is increasing evidence that the human brain might be organised in such hierarchical fashion (Essen & Maunsell, 1983; Fuster, 2004; Koechlin, Ody, & Kouneiher, 2003;) and evidence for such hierarchical arrangements have been found, for example, in rapid movement sequences such as keypresses with the fingers (Rosenbaum, Kenny, & Derr, 1983).

### 2.2. From primitives to composite inverse models

When it comes to the representation of human and robot actions, there are two aspects to hierarchical organisation: a somatotopic one and a functional one. A node at a certain level in the hierarchy has to assemble a motor plan deciding which body components are to be used, as well as how they are going to move, individually and in coordination to each other. Options include having the underlying components arranged sequentially (Demiris & Hayes, 2002), or additionally allow parallel execution (Demiris & Johnson, 2003). In the second case, an arbitration mechanism is needed that checks whether the underlying components running in parallel have any overlapping controlled degrees of freedom, using the somatotopic representation. There are no cross-inverse model constraints, such as speed of execution, for example.

Hierarchical *organisation* of motor structures (for selecting effectors and checking consistency) should be distinguished from their use during execution. Although structures can be arranged in the different ways described above, there are distinct possibilities regarding the use of the hierarchy during execution:

- Hierarchical structures are used during the planning phase, but once the sequence of commands have been planned (for example, by calling the lower levels with common parameterizations), and the higher levels have been populated, there is no further communication or online adjustment during execution between the different levels.
- There is an active recruitment and adjustment during the execution phase, where execution is passed to the lower levels; higher levels are coordinating them, for example, by determining their start and end times.

Although interesting data regarding the organisation of the mirror system have been reported (Rizzolatti et al., 1996; Umilta et al., 2001) there is little information regarding the temporal properties of the mirror system. We will review some recent evidence for this temporal dimension next.

### 2.3. Temporal coupling between action observation and execution

While fMRI and PET studies have demonstrated the existence of a mirror system in humans (Grezes et al., 2003), the temporal resolution limitations of brain scanning technology means that we are less knowledgeable regarding the temporal aspects of the mirror system. Recent experiments with transcranial magnetic stimulation (TMS) have, however, shed some light into the temporal coupling between action observation and execution (Gangitano, Mottaghy, & Pascual-Leone, 2001; Gangitano et al., 2004). Previous computational models of the mirror system have shown that such temporal coupling is crucial (Demiris, 1999; Demiris & Hayes, 2002; Oztop & Arbib, 2002). For example Demiris (1999), derived a set of testable predictions, most important of which was that monkey mirror neurons would not fire (or fire less) when the demonstrated movement was performed at speeds unattainable by the observer monkey. In Gangitano et al. (2001), it was shown that the amplitude of the motor evoked potentials (MEP) induced by TMS in humans observing a reaching-grasping action was modulated by the amount and timing of the observed finger aperture. A strict temporal coupling between corticospinal excitability and the dynamics of the reaching and grasping movement when passively observed was clearly demonstrated (Gangitano et al., 2001). A followup study (Gangitano et al., 2004) shed further light into the temporal characteristics of this coupling. The modulation in corticospinal excitability profiles during the observation of reach and grasp actions was studied under three experimental visual stimuli:

- Observation of natural reaching and grasping (RnG) actions.
- Observation of a RnG action where the appearance of the maximal finger aperture was significantly delayed.
- Observation of a RnG action where an unexpected finger closing and opening action was inserted before the final grasp portion of the demonstration.