
Proceedings Paper:
Buxton, D.R. orcid.org/0000-0001-5735-5927, Kerdegari, H., Mokaram, S. et al. (2 more authors) (2019) A window into the robot 'mind' : using a graphical real-time display to provide transparency of function in a brain-based robot. In: Martinez-Hernandez, U., Vouloutsi, V., Mura, A., Mangan, M. and Asada, M., (eds.) Biomimetic and Biohybrid Systems. 8th International Conference, Living Machines 2019, 09-12 Jul 2019, Nara, Japan. Springer International Publishing , pp. 316-320. ISBN 9783030247409

https://doi.org/10.1007/978-3-030-24741-6_28

A window into the robot 'mind': Using a graphical real-time display to provide transparency of function in a brain-based robot

David R. Buxton[1], Hamideh Kerdegari[2], Saeid Mokaram[3], Ben Mitchinson[1], and Tony J. Prescott[1]

[1] Adaptive Behaviour Research Group, The University of Sheffield, UK
http://abrg.group.shef.ac.uk
[2] Kingston University London, Kingston upon Thames, UK
[3] Speech and Hearing Research Group, The University of Sheffield, UK
http://spandh.dcs.shef.ac.uk

Abstract. Biomimetic robots are often given a humanoid or animaloid form that generates useful interaction affordances through similarities to natural counterparts. This has raised concerns about the potential for deception by creating the expectation of human- or animal-like intentional states that cannot (supposedly) be generated in artefacts. Here we report on the design of a graphical user interface (GUI) to the brain-based control system of the MiRo animal-like robot that we are developing to test the value of real-time displays as a means of increasing transparency for biomimetic robots and as a tool for investigating people's understanding of the relationship between internal mental processes and behaviour.

Keywords: biomimetic robot · brain-based robot · intentionality · MiRo robot · transparency

1 Introduction

Research in robot ethics has highlighted a potential trade-off between the utility of robots in human–robot interaction settings and their functional transparency. For instance, whilst robot developers such as Breazeal and Scassellati [1] have argued that "to interact socially, a robot must convey intentionality, that is, the human mind must believe that robot has beliefs, desires and intentions", Wortham and Theodorou [2] have proposed that present-day social robots may be effective only because they instil a belief in human users about intentional states that they do not actually have. In other words, robots might serve as effective social others by deceptively concealing the reality that they are machines controlled by computer programs. This risk of deception has been highlighted by a growing number of authors; Sparrow and Sparrow [3] have described social robots as intrinsically unethical, whilst the EPSRC "Principles of Robotics" [4], developed by a panel of UK ethicists and roboticists, advocates that as manufactured artefacts, robots "should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent".

The notion of robot deception has proved controversial [5] and further research is needed to understand the complex relationship between utility, transparency, and deception from ethical, philosophical, and psychological standpoints. More pragmatically, it will also be helpful to understand if robots can be useful in roles where they act as social others while still remaining transparent about their internal states and processes.

Wortham and Theodorou [2] have suggested various means of increasing the functional transparency of robots to explore these issues. For instance, by generating real-time audio or textual reports of the robot's control processes [6], and/or by using graphical real-time visualisation of the robot's inner workings [7], it may be possible to help human users construct a mental model (or 'Theory of Mind') more appropriate than the erroneous one that might otherwise develop through our human tendency to anthropomorphise.

In this paper, we report on the development of a graphical user interface (GUI) for the brain-based control system of the MiRo robot that we are developing to test the value of graphical real-time displays as a means of increasing transparency. Transparency is often discussed based on the presumption that the control systems underlying robot behaviour are fundamentally different from those operating through the nervous systems of animals to generate natural behaviour. MiRo presents an interesting case study in this context, as it is controlled by a highly simplified abstraction of the control architecture of the mammalian brain [8]. In addition to demonstrating how MiRo's behaviour is controlled, this GUI could therefore also serve as an educational tool to demonstrate how brains control bodies to generate behaviour. One potential outcome is that by seeing animal-like behaviour generated by a model of the mammalian nervous system, people may develop a more mechanistic view of the internal processes underlying their own thoughts and actions.

2 GUI overview

The MiRo graphical interface[†] (Fig. 1) is a hybrid display of dynamically updating and static information about MiRo's cognitive architecture, incoming sensory data, and internal computations that effectively represent the robot's current 'mental state'. Live data from sensory, attentional, affective, and action selection systems are presented in appropriately formatted plots and laid out in the form of an extended box–and–arrow diagram that guides understanding of how these components interact to drive behaviour. This creates a visualisation of the complete 'brainstem' level of MiRo's hierarchical cognitive architecture [8] and of information that would otherwise remain entirely hidden. Many of the dynamic plots further invite the user to click through to an enhanced view that provides access to more detailed information or explains the component in greater depth.

[†] Available at: https://github.com/hamidehkerdegari/graphical_interface

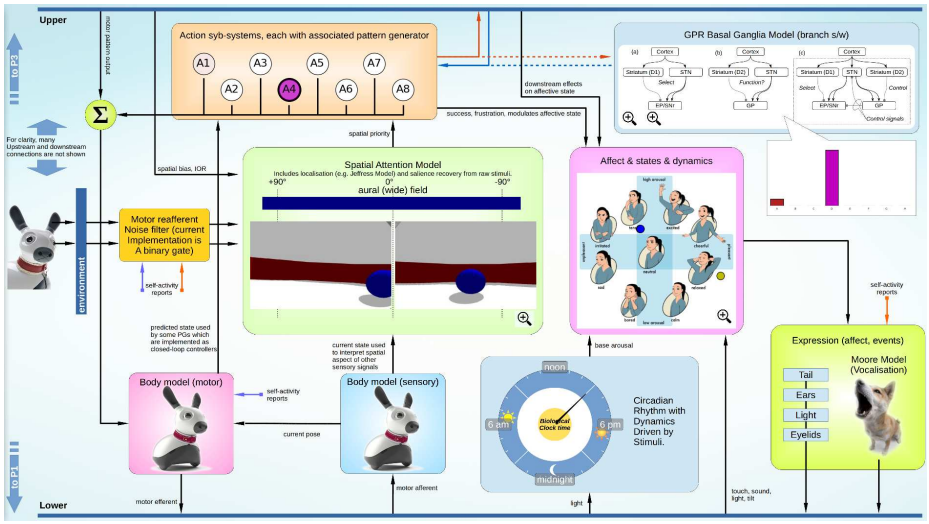

Fig. 1. Main window of the MiRo graphical interface, showing an overview of a virtual robot's cognitive state while approaching a toy ball. Component subsystems are framed in different colours; action selection is orange, spatial attention is green, circadian clock is pale blue, and affect is magenta.

2.1 Component summaries

Action selection: Displays the current input salience of each action subsystem and the corresponding level of disinhibitory output from the basal ganglia model, illustrating the important point that even when several or no high-salience inputs exist, a selection system should still select a single action strongly and unambiguously [9].

Spatial attention: Displays MiRo's visual field and aural attention indicator[‡], with an enhanced view (Fig. 2) that includes MiRo's visual salience map.

Circadian clock: MiRo's internal circadian clock, which impacts affective state and drives a periodic sleep cycle.

Affect: Shows emotion, mood, and sleepiness, which are all represented by a 2D circumplex model [10] influenced by sensory and cognitive factors.

3 Conclusion

The graphical interface has several potential benefits. Firstly, the information presented may prove useful as a STEM teaching resource; the diagrammatic representation of a model cognitive architecture provides a visual aid that illustrates

[‡] The virtual MiRo environment does not support auditory simulation, therefore the spatial attention component shown here lacks that information.

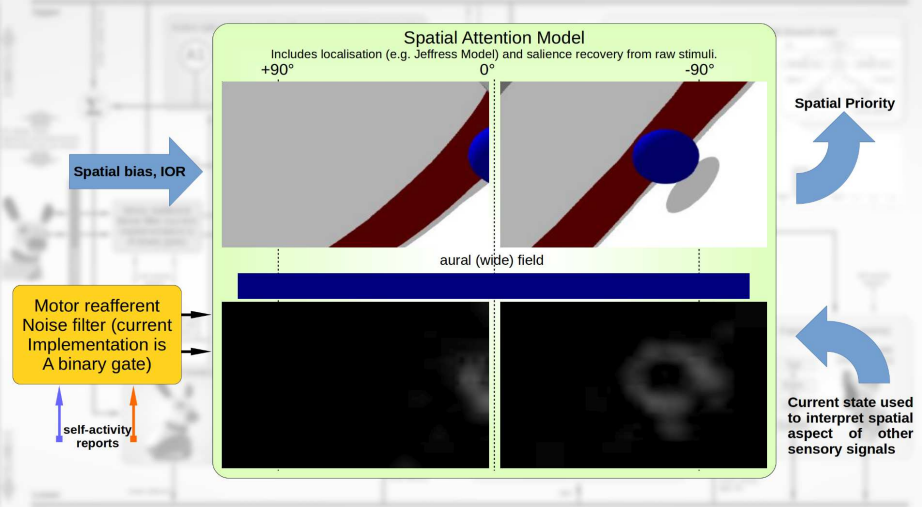

Fig. 2. Enhanced view of the spatial attention component showing MiRo's visual field *(upper)* and salience map *(lower)*. Objects attracting attention, such as MiRo's toy ball, will be highlighted in the salience map.

the functional connectivity of mammalian brains, and the integration of live sensory information may facilitate discussions on the problem of action selection and the role of specific brain structures. Secondly, the GUI greatly increases the transparency of MiRo's behaviour, helping to clarify the similarities and differences between MiRo's brain and human brains, to explain the functionality underlying MiRo's behaviour, and to refute beliefs that MiRo is truly conscious.

Furthermore, because the interface displays live information directly from the active 'mind' of a behaving robot, it is also interactive; not only can a user study the diagram to learn how spatial attention drives action selection and sensory stimulation modulates affect, but as they interact with MiRo they can observe the robot attending to their movements, choosing to approach them, and the increase in affect that underlies the wagging tail if they pet him.

We are interested in exploring how such operational transparency may improve human–robot relationships [2], and we are optimistic that this GUI presents a valuable opportunity to deepen the public's interest in biomimetic robots. We plan to utilise the MiRo GUI in future experimental work to explore if the benefits described here are realised in practice, and how it influences users' understanding not only of our simulated cognitive architecture, but also of their own.

References

[1] C. Breazeal and B. Scassellati, "How to build robots that make friends and influence people," in *Proceedings 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human and Environment Friendly Robots with High Intelligence and Emotional Quotients (Cat. No.99CH36289)*, vol. 2, Oct. 1999, pp. 858–863.

[2] R. H. Wortham and A. Theodorou, "Robot transparency, trust and utility," *Connection Science*, vol. 29, no. 3, pp. 242–248, Jul. 3, 2017.

[3] R. Sparrow and L. Sparrow, "In the hands of machines? the future of aged care," *Minds and Machines*, vol. 16, no. 2, pp. 141–161, 2006.

[4] M. Boden, J. Bryson, D. Caldwell, K. Dautenhahn, L. Edwards, S. Kember, P. Newman, V. Parry, G. Pegman, T. Rodden, T. Sorrell, M. Wallis, B. Whitby, and A. Winfield, "Principles of robotics: Regulating robots in the real world," *Connection Science*, vol. 29, no. 2, pp. 124–129, 2017.

[5] E. C. Collins, "Vulnerable users: Deceptive robotics," *Connection Science*, vol. 29, no. 3, pp. 223–229, 2017.

[6] R. H. Wortham and V. Rogers, "The muttering robot: Improving robot transparency though vocalisation of reactive plan execution," in *26th IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man) Workshop on Agent Transparency for Human-Autonomy Teaming Effectiveness (Lisbon:)*, 2017.

[7] R. H. Wortham, A. Theodorou, and J. J. Bryson, "Improving robot transparency: Real-time visualisation of robot ai substantially improves understanding in naive observers," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 1424–1431.

[8] B. Mitchinson and T. J. Prescott, "MIRO: A robot "mammal" with a biomimetic brain-based control system," in *Biomimetic and Biohybrid Systems*, N. F. Lepora, A. Mura, M. Mangan, P. F. Verschure, M. Desmulliez, and T. J. Prescott, Eds., vol. 9793, Cham: Springer International Publishing, 2016, pp. 179–191.

[9] T. J. Prescott, P. Redgrave, and K. Gurney, "Layered control architectures in robots and vertebrates," *Adaptive Behavior*, vol. 7, no. 1, pp. 99–127, Jan. 1999.

[10] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and Psychopathology*, vol. 17, no. 3, pp. 715–734, Sep. 2005.