Safety Control Architecture for Personal Robots: Behavioural Suppression with Deliberative Control

Roger Woodman, Alan Winfield, Chris Harper, Mike Fraser

***Abstract*—This paper presents a novel robot control architecture for use with personal robots, and argues its potential for improving the safety of these types of system, when compared to existing approaches. The proposed architecture design separates the control system into two distinct areas, one area responsible for safe operation and the other for coordinating tasks. The architecture design is formed in a hierarchical structure, composed of low-level deliberative control modules and high-level behavioural safety modules. It is argued that as a result of removing safety considerations from the design of task routines, increasingly complex tasks can be completed safely, which are both more flexible to environmental changes and easier to coordinate.**

I. Introduction

Each decade since the 1970's it has been said that, in 10 years time people working with robots would be common place in both industry and at home. Although the question of why this hasn't happened yet is being continually asked, the answer has changed over the years. With the development of new engineering techniques, miniaturisation of electronics and the increase of computing power, designers are now at the stage where a robot system is capable of performing useful cooperative tasks with human users. However, the problem which designers now face, and for which the deployment of robot systems is being impeded, is that of safety. Safety implications have always been a concern for robot designers and traditionally the solution has been to prevent the user coming into contact with the robot, by means of physical barriers. For personal robots to become a reality these barriers will need to be removed and more dynamic flexible safety methods introduced.

In this paper we present an investigation into different types of robot control architectures and propose a new architecture model, which aims to address a number of key deficiencies found in other architecture types. The architecture design focuses on safety and aims to separate the task control of the robot from control routines associated with safety. The motivation for this design is the premise that safety control should be treated as a separate constant process, running in parallel with other robot activities. It is argued that in this way the robot control designer can appropriately separate reactive and deliberative components of the controller design, which can be a difficulty of existing architecture types.

Manuscript received May 28, 2010.

E-mail addresses: roger.woodman@brl.ac.uk (R. Woodman), alan.winfield@uwe.ac.uk (A. Winfield), cjharper@avian-technologies.co.uk (C. Harper), fraser@compsci.bristol.ac.uk (M. Fraser).

A. Background

Early research [1] into the safety of personal robots suggested that industrial robots could be adapted and used to perform human-robot interactive (HRI) tasks. However, as Alami et al. [2] discusses, it is not feasible to take a large rigid robot and adapt it to the delicate tasks necessary for a personal robot. At present, the main method for making robots safe is to prevent people from stepping into the working area of the robot. This approach, developed by manufacturing industries, maintains a clear space around the robot. Access is prevented by means of physical barriers and proximity sensors, which halt the robot on activation [3]. It is apparent that if robots are to interact directly with humans, a new approach to safety is needed.

All industries, which require the development of safety-critical systems, have strict processes and standards which must be followed before the system can be put into service. However, as discussed by Desantis et al. [4] and Kulic and Croft [5], there are still no safety standards for complex robots for use in cooperative situations with humans.

One objective of robot design, which every engineering designer strives to achieve, is that of intrinsic safety. Intrinsic safety is the property that a system cannot inherently cause a hazard, even if it fails or malfunctions. In robotics there are a number of well established intrinsic safety techniques [6], [7]:

- Actuators with limited power/speed which guarantee safe behaviour in case of a fault
- De-energised brakes, which halt the robot in the event of a power failure
- The use of a 'dead man's switch', which must be engaged in order for the robot to operate

Although an intrinsically safe design is what designers aim to achieve, it can be difficult to prove this for a complex system by means of testing. Therefore, designs which are said to be intrinsically safe, are generally functionally and physically simple.

Research by Marzwell [8], reveals two classes of potential hazards that can exist in robot controllers. These are 'system level' failures, caused by the controller itself, and 'task level' failures, which are caused by valid commands to the controller that result in an unsafe event i.e. collision, unbalancing or other hazards. To alleviate some of the issues associated with traditional controller designs, a number of new approaches have been developed. The most popular of which, is to modularise a system into a group of inter-

connected units that can be developed and tested in isolation. As Laibinis and Troubitsyna [9] identify, "Traditionally abstraction, modularisation and layered architecture are recognised to be effective ways to manage system complexity".

The remainder of the paper will focus on modular hierarchical based robot control architectures.

B. Related Work

The latest research into the safety of personal robots [10], [2], [7] argues that it is not possible to control a complex robot system in a dynamic environment, using traditional control methods. Instead they suggest a behavioural type system, which can react and adapt to changing conditions.

Research by Bensalem et al. [10] and Lussier et al. [11], has shown that a hierarchical approach to system safety, with different control layers providing planning, task execution and safety supervision, can improve the reliability and dependability of an autonomous robot system. This research is broadly based on the behaviour-based techniques developed by Rodney Brooks. In Brooks' work [12], [13], he demonstrated how different simple behaviours could be combined to produce new complex behaviours. Both the Subsumption type architecture and 'three layer' control architecture can be seen in figure 1.1.

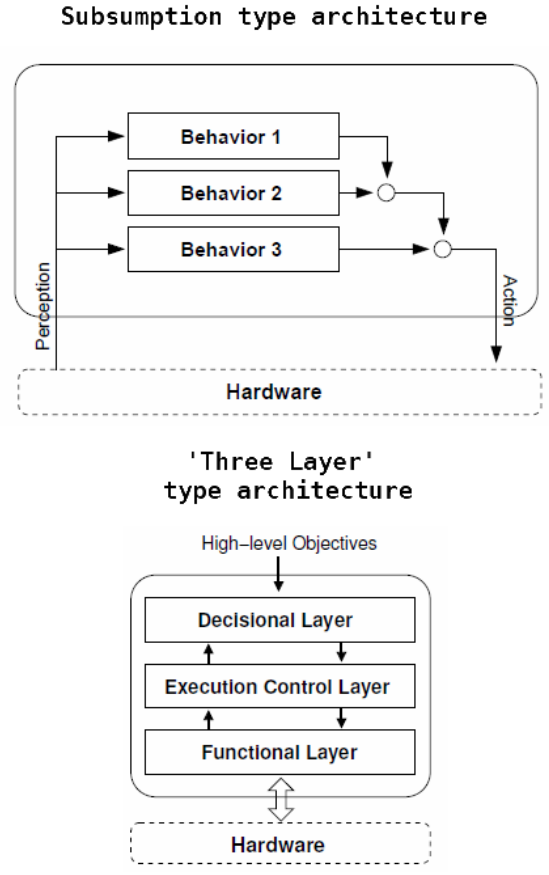

Figure 1.1: Subsumption architecture and 'three layer' control architecture [14].

A notable architecture using the three layer approach is the LAAS architecture shown in figure 1.2. This architecture divides the software controlling the robot into three levels: decisional; execution; functional. The distinct feature of the LAAS architecture is the functional level, which encapsulates groups of sensors into modules which can communicate with other modules via a service link.

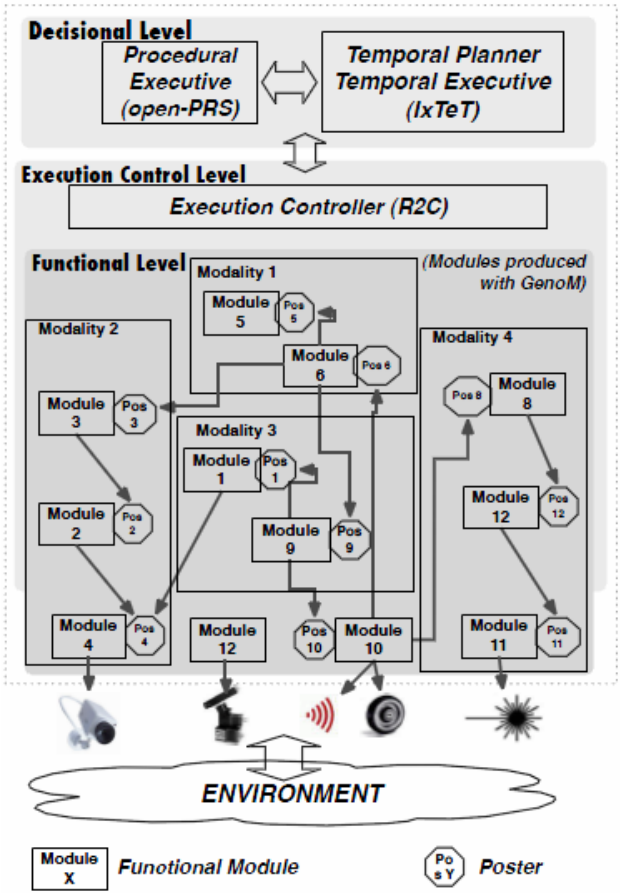

Figure 1.2: LAAS architecture [14].

The abstraction of the functional elements of the robot control software, allows for modules to be added, removed and amended while retaining control functionality of the robot.

The research of this paper seeks to build on the research into hierarchical modular control architectures and adapt it for use as a safety control architecture for HRI robots. This will be expanded on in the following sections.

C. The Research

This paper presents initial research into a new type of robot control architecture design. The aim of which is to produce a dynamic control system capable of adapting control routines to maintain safety, while continuing to perform useful tasks. The architecture design has been developed by analysing the strengths and weaknesses of a number of other control systems.

This research will focus on the safety of human-robot

interaction with a class of robots that pose a risk based on their physical size, strength and behaviour. Although it is noted that there are safety concerns with smaller robots, such as the iRobot Roomba vacuum cleaning robot. The risks associated with these types of robot are such that they can be developed using the same safety criteria as other small electrical devices [6], [2].

The intended application for this research, is for robot systems that are used in both industry and the home, where the robot would be working on cooperative tasks with a human user. These tasks could be anything from, cooking, building flat-pack furniture or helping the stocking of shop shelves. Particular focus will be made on multi-functional robots, which can perform a variety of tasks, as it is asserted that these types of robots would have all the safety consideration of task-specific robots with additional considerations due to their generalised design.

D. Research Questions

Based on the review of current research and preliminary experimentations, the following questions have been formulated:

- To what extent must individual behaviour modules be adapted, in order to maintain safety in the face of changing environments and/or machine dynamics?
- By purely suppressing perceivably unsafe control actions, is it possible to complete a task while avoiding hazards?
- Is it possible to design a robotic system based on separate safety and task modules, where the task modules can be changed and the safety modules remain unaltered and the safety uncompromised?
- To what extent do the relative priorities of tasks need to be changed, in order to maintain safety in HRI tasks?
- Can a robotic system be designed which can learn how to complete tasks safely?

II. Robot Control Architectures

Robot control architectures can be broadly divided into one of three categories: deliberative; reactive and hybrid [15]. A deliberative controller uses the sense-plan-act method for completing tasks. This involves reasoning about the perceived world and acting appropriately. This contrasts to reactive controllers, which employ a sense-act approach. This approach avoids processing and storage overheads, often associated with reasoning about the state of the world, both internal and external to the robot. The final type of robot controller, the hybrid controller, is a combination of both the deliberative and reactive controller types.

Many argue that a hybrid controller type provides the only way to control a robot performing complex tasks in a dynamic environment [16], [17]. As Bonasso and Kortenkamp [18] discusses, a control architecture is needed which can accept new tasks and information and react to the world at any time; "We do not want an architecture that requires a robot to be reprogrammed each time its goal changes." [18]. Equally robot controllers that are purely reactive are not able to complete complex tasks which require coordination and planning. As Martin Proetzsch [16] states, "the problem of controlling complex robotic systems is not solved by the behaviour-based paradigm alone. Rather, while helping with some common problems, behaviour-based architectures introduce new difficulties.".

A. Hybrid Architectures

Hybrid robot control architectures generally divide the controlling task into reactive and deliberative modules, with a communication layer in-between to organise control events. The diagram in figure 2.1 is a typical representation of a hybrid control architecture.

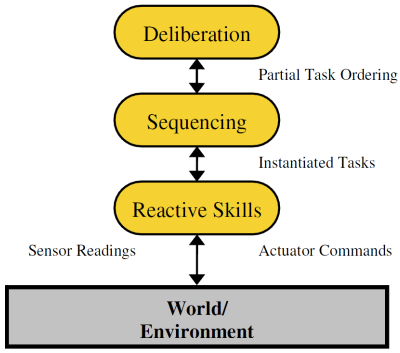

Figure 2.1: 3 tier control architecture [19].

The majority of robot control architectures [10], [17], [18], which have separate safety modules, use deliberative control modules for completing tasks and reactive safety modules to monitor the behaviour of the robot and prevent any unsafe actions. An alternative style architecture is the Sensor Fusion Effects (SFX) architecture, developed by Murphy and Arkin [20]. This uses a layered approach, with a low-level behavioural task layer, providing the functionality, and a top-level deliberative safety layer, which monitors the actions of the robot and prevents any unsafe events. As Murphy [21] discusses, one benefit of a reasoning safety layer, is that the task of the controller can be readily customised without impacting predefined safety routines.

III. Our Research – Architecture Design

The research being discussed in this paper aims to develop a new safety architecture, which takes the reactive elements from behavioural systems and the procedural elements from traditional control systems. This will be used to create a new method for designing safety critical human-robot interactive systems, that can be proven and ultimately certified as safe.

By analysing the existing robot control architectures we have identified a number of strengths and weaknesses and have formulated a criteria based on these findings. The

following list of design considerations summarise these criteria:

- Hierarchical structure, with separate safety and control modules
- Safety modules designed to maximise re-usability and minimise re-testing
- Safety modules should be independent of task modules, allowing for the task layer to be changed, without compromising the safety of the robot
- Compatibility with any controller type, i.e. PID, neural network, fuzzy logic
- Appropriate for any robot hardware comprised of sensors and actuator

A. High-level Diagram

The motivation behind the design shown in figure 3.1, is based on a number of identified problems with other similar architecture types. The design has also been developed to both maximise the interaction between the individual layers and provide sensor data to all layers simultaneously. The main characteristics of the design are as follows:

- Safety and task layers are separated
- Multiple safety and task layers
- Safety layer has no dependence on the task layer
- All layers have access to sensor data
- Hierarchical structure, with safety layers having highest priority
- A lower priority layer cannot affect a higher priority layer.
- Ability to suppress actions of lower layers (behaviour arbitration)
- Layers are prevented from altering data or injecting commands into other layers
- Event information is passed down to lower layers (information such as suppression details, warnings, advisory details)
- Event information is used for learning and adaptation

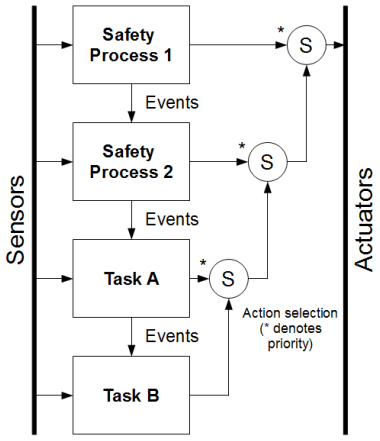

Figure 3.1: Safety control architecture.

It is important to reiterate that although many of these design consideration are similar to the Brooks' Subsumption Architecture [12], the ability for lower layers to change upper layers will not be allowed. This is due to the assertion that if a safety layer was proven to be safe, then allowing lower layers to make changes would invalidate the overall safety of that particular module.

B. Safety Layer

The safety layer will be made up of a number of behavioural safety modules. An example of how safety policies will be implemented is shown in figure 3.2. In this example there are two modules, which implement safety policies to monitor the robot arm, wheel and gripper. As the diagram illustrates, both safety modules monitor the gripper, therefore, if either (but not both) modules fail, safety is maintained for the gripper. In addition, as the sensor input of each module is slightly different, a number of sensors (cam 2, sonar 1 or sonar 2) can fail, without compromising the safety of the gripper.

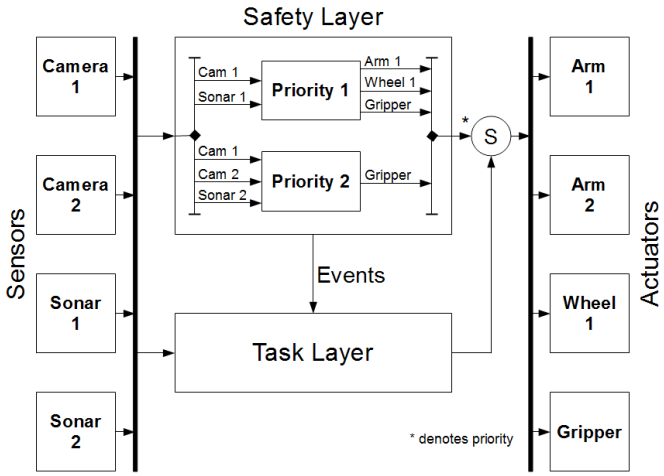

Figure 3.2: Safety control architecture - Safety modules with redundancy.

This design has two main benefits. The first is that safety modules/policies can be designed for a sensor and actuator combination, regardless of the type of system (robot or otherwise), and then added to another system as a pre-tested and proven safety unit. The second benefit, and the reason behind using a module approach, is the ease of which redundancy can be added to a system, by simply adding additional safety modules.

C. Task Layer

The task layer will implement all of the control routines needed to plan, organise and complete tasks. This means that unlike the safety layer, which is purely behavioural, the task layer must be deliberative, in order to handle the complex tasks required for a personal robot. At present the task layer specification is under development. Although we have many ideas on how tasks could be dealt with, experimentation is required in order to make sense of this collection of ideas.

Our current line of investigation is looking at the possibility of taking an existing robot controller and wrapping a safety layer around it, in order to produce a new 'safe' controller. This would fit the current architecture design, as the existing controller could be integrated into the task layer with minimal changes. This approach would allow designers to take the best parts of traditional control system design and combine it with the flexibility and reactive nature of a behavioural safety system.

IV. Future Work

We are currently working on a robot simulation tool which, when complete, will allow us to perform accurate consistent experiments with a number of robot architecture types. With this tool we believe we will be able to rapidly develop the architecture design and identify any shortcomings.

A. Further Development to Architecture Design

The subject of this research paper is a relatively under-studied area, therefore there are many directions this research could take. At present we are investigating techniques for action planning and sequencing. This investigation is currently looking at methods proposed by Hertzberg et al. [22] and Arkin and Balch [23], with focus particularly on deliberative sequencers, which use task information to decide the order that control actions should be executed.

B. Proving Design – Experiments

The work presented in this paper is part of a larger research project. The overall goal of this research project is to develop a new safety architecture, which can be demonstrated to improve the safety, reliability and dependability of a complex robot system performing human-robot interactive tasks.

In order to test and evaluate the development of each safety architecture design, it is essential that real world tasks are performed. The design of these tasks will cover a wide range of hazardous issues that may arise while working in proximity to a robot. Such hazards, which must be accounted for include:

Direct Hazards

- Collision with human user
- Collision with surrounding objects
- Collision with other humans (non-user)
- Collision with robot body

Indirect Hazards

- Dropping an object
- Causing a human to move into a dangerous situation
- Spillage while moving an object
- Burning caused by prolonged application of heat i.e. while ironing

V. Conclusion

This paper has presented a novel robot controller architecture design, which emphasises safety, and demonstrates how a control system can be developed with separate safety and task processes. The argument put forward by this paper is that a hierarchical control architecture, composed of low-level deliberative control modules and high-level behavioural safety modules, can be used to greatly reduce the safety aspects of the controller design. This abstraction of safety and control, allows designers to continually develop and update the controller design, with fewer implications to the overall safety of the robot, when compared to a traditional controller that has no distinction between control and safety.

As discussed, this initial research will be followed by a series of experiments. These experiments will be used to both improve the architecture design and to make quantitative comparisons with other robot controllers.

References

[1] Kulić, D. and E. Croft (2006), "Real-time safety for human-robot interaction." *Robotics and Autonomous Systems*, 54(1), 1-12.

[2] Alami, R., A. Albu-Schaeffer, A. Bicchi, R. Bischoff, R. Chatila, A. De Luca, A. De Santis, G. Giralt, J. Guiochet, G. Hirzinger, F. Ingrand, V. Lippiello, R. Mattone, D. Powell, S. Sen, B. Siciliano, G. Tonietti, and L. Villani (2006), "Safe and dependable physical human-robot interaction in anthropic domains: State of the art and challenges." *Proc. IROS'06 Workshop on pHRI - Physical Human-Robot Interaction in Anthropic Domains.*

[3] Ogorodnikova, O. (2008), "Methodology of safety for a human robot interaction designing stage." *2008 Conference on Human System Interaction*, 452-457.

[4] Desantis, A., B. Siciliano, A. Deluca, and A. Bicchi (2008), "An atlas of physical human robot interaction." *Mechanism and Machine Theory*, 43, 253-270.

[5] Kulić, D. and E. Croft (2003), "Strategies for safety in human-robot interaction." *Proc. IEEE Int. Conf. on Advanced Robotics*, 810-815.

[6] Wyrobek, Keenan A., Eric H. Berger, H. F. Machiel Van der Loos, and J. Kenneth Salisbury (2008), "Towards a personal robotics development platform: Rationale and design of an intrinsically safe personal robot." *ICRA*, 2165-2170. 38.

[7] E. Dombre, F. Pierrot G. Duchemin L. Urbain, Ph. Poignet (2001), "Intrinsically safe active robotic systems for medical applications." *Proceedings of IARP/IEEE-RAS Joint Workshop on Technical Challenge for Dependable Robots in Human Environment.*

[8] N. Marzwell, M. Hecht, K.S. Tso (1994), "An integrated fault tolerant robotic control system for high reliability and safety." *Proceedings of Technology 2004*.

[9] Laibinis, Linas and Elena Troubitsyna (2005), "Formal development of reactive fault tolerant systems." *2nd Workshop on Rapid Interaction of Software Engineering Techniques.*

[10] Bensalem, Saddek, Matthieu Gallien, Felix Ingrand, Imen Kahloul, and Thanh-Hung Nguyen (2009), "Toward a more dependable software architecture for autonomous robots." *Special issue on Software Engineering for Robotics of the IEEE Robotics and Automation Magazine,* 16, 67-77.

[11] Lussier, Benjamin, Matthieu Gallien, Jrmie Guiochet, Flix Ingr, Marc olivier Killijian, and David Powell (2007), "Experiments with diversified models for fault-tolerant planning." *IARP'07*, Roma, Italy.

[12] Brooks, Rodney A. (1999), *Cambrian intelligence: the early history of the new AI*. MIT Press, Cambridge, MA, USA.

[13] Toal, Daniel, Colin Flanagan, C. Jones, Bob Strunz, Daniel Toal, Colin Flanagan, Caimin Jones, and Bob Strunz (1996), "Subsumption architecture for the control of robots." *IMC-13*, 703-711.

[14] Lussier, B., R. Chatila, F. Ingrand, M.O. Killijian, and D. Powell (2004), "On fault tolerance and robustness in autonomous systems." *3rd IARP-IEEE/RAS-EURON Joint Workshop on Technical Challenges for Dependable Robots in Human Environments.*

[15] Bryson, Joanna (2002), "Hierarchy and sequence vs. full parallelism in action selection." *In Proceedings of the Sixth Intl. Conf. on Simulation of Adaptive Behaviour*, 147-156.

[16] Martin Proetzsch, Karsten Berns, Tobias Luksch (2010), "Development of complex robotic systems using the behavior-based control architecture iB2C." *Robotics and Autonomous Systems*, 58.

[17] Nordin, Peter and Mats Nordahl (1999), "An evolutionary architecture for a humanoid robot." *Proceedings 4th International Symposium on Artificial Life and Robotics.*

[18] Bonasso, R. Peter and David Kortenkamp (1996), "Using a layered control architecture to alleviate planning with incomplete information." *Proceedings of the AAA Spring Symposium on Planning with Incomplete Information for Robot Problems*, 1-4.

[19] Narayan, Pritesh P., Paul P.Y. Wu, Duncan A. Campbell, and Rodney A. Walker (2007), "An intelligent control architecture for unmanned aerial systems (UAS) in the national airspace system (NAS)." *2nd International Unmanned Air Vehicle Systems Conference.*

[20] Murphy, R. R. and R. C. Arkin (1992), "Sfx: An architecture for action-oriented sensor fusion." *IROS '92 - Proceedings of the 1992 IEEE RSJ International Conference on Intelligent Robots and Systems,* 1079-1086.

[21] Murphy, Robin R. (1997), "Intelligent sensor fusion for the 1997 AAAI mobile robot competition." *AAAI*, 795-796.

[22] Joachim Hertzberg, Frank Schönherr, Schloß Birlinghoven (2001), "Concurrency in the DD&P Robot Control Architecture." *Proc. of The Int. NAISO Congress on Information Science Innovations (ISI'2001)*.

[23] Arkin, Ronald C. and Balch, Tucker (1997), "AuRA: principles and practice in review." *Journal of Experimental & Theoretical Artificial Intelligence*, 9:2, 175-189.