We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

12

From Robot Arm to Intentional Agent: The Articulated Head

Christian Kroos, Damith C. Herath and Stelarc
MARCS Laboratories, University of Western Sydney
Australia

1. Introduction

1.1 Robots working together with humans

Robot arms have come a long way from the humble beginnings of the first Unimate robot at a General Motors plant installed to unload parts from a die-casting machine to the flexible and versatile tool ubiquitous and indispensable in many fields of industrial production nowadays. The other chapters of this book attest to the progress in the field and the plenitude of applications of robot arms. It is still fair, however, to say that currently industrial robot arms are primarily applied in continuously repeated manufacturing task for which they are pre-programmed. They are known for their precision and reliability but in general use only limited sensory input and the changes in the execution of their task due to varying environmental factors are minimal. If one was to compare a robot arm with an animal, even a very simple one, this property of robot arm applications would immediately stand out as one of the most striking differences. Living organisms must sense changes in the environment that are crucial to their survival and must have some flexibility to adjust their behaviour. In most robot arm contexts, such a comparison is currently at best of academic interest, though it might gain relevance very quickly in the future if robot arms are to be used to assist humans to a larger extend than at present. If robot arms will work in close proximity with and directly supporting humans in accomplishing a task, it becomes inevitable for the control system of the robot to have far reaching situational awareness and the capability to adjust its 'behaviour' according to the acquired situational information. In addition, robot perception and action have to conform a large degree to the expectations of the human co-worker.
Countless situations can be imagined (and are only a step away from current reality while fully autonomous mobile robots might still be far off):

- A robot arm lifting and turning a heavy workpiece such as a car engine for human inspection and repair;
- A robot arm acting as a 'third hand' for a human worker for all kinds of construction and manufacturing work that is yet too complex to be fully automated;
- A robot arm assisting a temporarily or permanently bedridden person and/or the nurses taking caring of the person. For the latter, one of the most important tasks would be again the careful lifting of the person;
- An intelligent robotic device assisting people with walking difficulties replacing the current clunky walkers;

- A robot arm assisting elderly people at home with all tasks that require considerable force (from opening a jar to lifting heavy items) or involve difficult to reach places (which might be simply the room floor).

To assess the social and economical impact that such a development would have, one might draw a parallel to the revolution that heavy machinery meant for construction and agriculture and with this for society at large. Within one generation, one might speculate, it could become inconceivable to imagine many workplaces and the average home in industrialised countries without assisting robot arms.

1.2 Joint action

Humans collaborate frequently with each other on all kinds of tasks, from jointly preparing a meal to build a shelter to write a book about robot arms. Even if the task is very simple such as carrying a load together, the underlying coordination mechanism are not. Collaborations with physical co-presence of the actors require a whole gamut of perceptive 'cues' to be observed and motor actions to be adjusted. This might be accomplished during execution or already during planning taking into account predictions of the co-workers' actions. In almost all situations so-called *joint attention* (to which we will return shortly) is an additional prerequisite. The emerging field of *joint action* research in psychology (Sebanz et al., 2006) tries to unravel the perceptive, cognitive and motor conditions and abilities that allow the seemingly effortless coordination of human action to accomplish a common goal. Sebanz et al. (2006) suggest an operational definition of joint action as 'any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment'. In this regard, the requirement for joint action builds on the concept of joint attention and extends it by requiring the prediction of actions of another. Joint action therefore depends on the abilities to (1) share representations, (2) predict actions, and (3) integrate predicted effects of one's own and the other's actions. These requirements do not change if the other is a machine or - narrowed down given the topic of this book - a robot arm. Admittedly, one could offload all the coordination work to the human co-worker by 'stereotyping' the action of the robot arm, i.e. reduce the movement vocabulary and make it easily predictable in all situations, but one would at the same time also severely limit the usefulness of the robot arm.

Arguably, we humans excel in joint actions because we perceive other humans as intentional agents similar to ourselves. Whether or not this would apply to robots is at the current state of research an unanswered question and, moreover, a question that poses difficulties to any investigation as there is no direct access to the states of the human mind. Some studies, though, provided partial evidence in favour of this using sophisticated experiment designs. Participants have been found to attribute animacy, agency, and intentionality to objects dependent on their motion pattern alone (Scholl & Tremoulet, 2000) and studies in Human-Robot Interaction (HRI) confirmed that robots are no exceptions (though clear differences remain if compared to the treatment of motor actions of other humans; see Castiello, 2003; Liepelt et al., 2010). Humans might also attribute emotions and moods to robots (e.g. Saerbeck & Bartneck, 2010). An important aspect of considering a robot as an intentional agent is the tacitly included assumption that the actions of the robot are neither random nor fully determined (as both would exclude agency), but a more or less appropriate and explainable response to the environment given the current agenda of the robotic agent.

Note that 'intentional agent' does not equate with human-like: animals are intentional agents as well, and there is long history of collaboration of humans with some of them, one of the

most perspicuous examples being shepherds and their dogs. While high-level understanding of conspecifics as intentional beings like the self (so called 'theory of mind', see Carruthers & Smith (1996) for a theoretical review) might be a cognitive competency that is limited to humans and maybe (Tomasello, 1999) - or maybe not (Call & Tomasello, 2008) - other primates, understanding others as intentional beings similar to oneself is not a capability that emerged ex nihilo. Over the last two decades, research concerned with the development of this capacity has indicated that it is closely tied to what is now generally called *joint attention* (Tomasello et al., 2005).

1.3 Joint attention

The concept of joint attention refers to a triadic relationship between two beings and an outside entity (e.g. an object like an apple) whereby the two beings have a shared attentional focus on the object. Joint attention has been seen as a corner stone in the development of social cognition and failure to achieve it has been implicated in Autistic Spectrum Disorders (Charman, 2003). As pointed out by Tomasello (1999), for joint attention to be truly joint, the attentional focus of the two beings must not only converge on the same object but both participants must also monitor the other's attention to the object (secondary inter-subjectivity). This should be kept in mind when thinking about a robot arm collaborating with humans as it basically requires some kind of indicator that the control system is aware of the current human actions and - at least potentially - is able to infer the intention of the human co-worker. This indicator might be a virtual or mechatronic pair of eyes or full face. In previous research on joint attention, a variety of different definitions have been used, not all of them as strict as Tomasello's. This is because applying his definition poses substantial difficulties in verifying whether joint attention has occurred in an experimental set-up, in particular when investigating infants or non-humans, and by extension also makes modelling it in a machine more difficult.

Its link to understanding other people as intentional beings notwithstanding, joint attention is not uniquely human; it has been observed in monkeys (Emery et al., 1997) and apes (Carpenter et al., 1995). In the latter study, joint attention was heuristically defined in terms of episodes of alternating looks from an ape to the person and then to the object. This way of quantifying joint attention through gaze switching has become the one most frequently used, even though gaze alternation is not always a reliable indicator of joint attention as mentioned above. Furthermore, gaze alternation constitutes neither a sufficient nor a necessary condition for joint attention. On the one hand, it is very common among animals to use another animal's gaze direction as a clue to indicate important objects or events in the environment but the fact that the other animal paid attention to this event is of no consequence and not understood (Tomasello, 1999); on the other hand establishing joint attention, for instance, through the use of language is a much more powerful mechanism than just gaze following (since it includes the aspect of the object or event on which to focus). All of this will have an impact on designing a robot arm control system that is able to seamlessly and successfully cooperate with a human. Not surprisingly, joint attention in robotics poses challenges not to be underestimated (Kaplan & Hafner, 2004).

2. A virtual agent steps into the physical world

We went into some details with regard to joint action and attention to explain some of the basic motivations driving our use of a robot arm and shaping the realisation of the final system, the Articulated Head. Because of its genesis as a work of art, many of our aims and many of

Fig. 1. The Articulated Head.

the properties of the Articulated Head are probably far beyond the ordinary in robot arm research and development. On the hardware side, the Articulated Head consist of a Fanuc LR Mate 200iC robot arm with an LCD monitor as its end effector (see Figure 1). The Articulated Head represents the robotic embodiment of the *Prosthetic Head* (Stelarc, 2003) by Australian performance artist Stelarc, an Embodied Conversational Agent (ECA) residing only in virtual reality, and is one of the many faces of the *Thinking Head* developed in the Thinking Head Project (Burnham et al., 2008).

The Prosthetic Head (Figure 2) is a computer graphic animation based on a 3D laser scan of the head of the artist. Through deforming its underlying 3D mesh structure and blending the associated texture maps a set of emotional face expressions and facial speech movements are created. A text-to-speech engine produces the acoustic speech output to which the face motion are synchronised. Language input from the user is acquired through a conventional computer keyboard. Questions and statements from the user are sent to the A.L.I.C.E. chatbot (Wallace, 2009) which generates a response utterance. The Prosthetic Head has been presented at numerous art exhibitions, usually as a projection of several square meters in size.

The Articulated Head was born as a challenge to the traditional embodiment of ECAs in virtual reality. No matter how convincing the behavioural aspects and cognitive capabilities

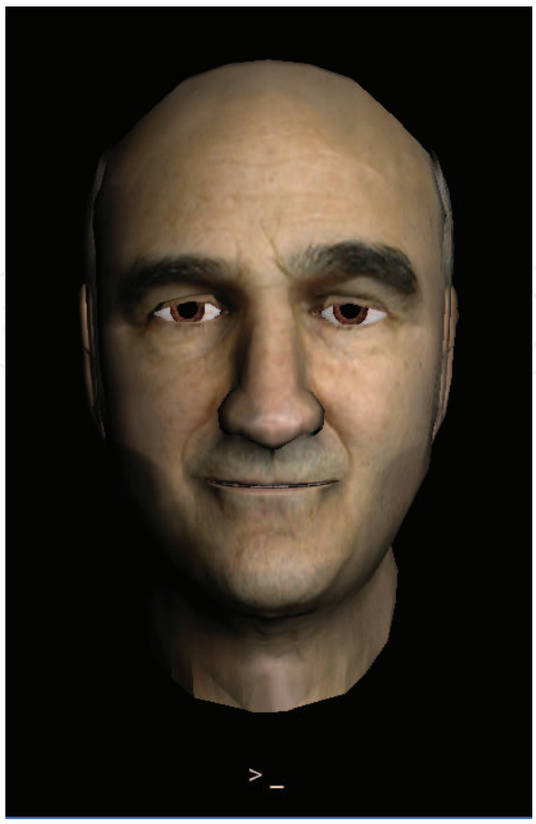


Fig. 2. The Prosthetic Head.

of a conventional ECA might be, it would always fall short of sharing the physical space with the interacting human. As physical co-presence is of great importance for humans (e.g. infants do not learn foreign language sounds from television; see Kuhl et al., 2003), transgressing the boundaries of virtual reality would enable a different quality of machine-human interaction. The robot arm enables the virtual agent to step out into the physical space shared with its human interlocutor. The sensory capabilities of the Articulated Head in the form of cameras, microphones, proximity sensors, etc. (Kroos et al., 2009) allow it to respond to the user's action in the physical world and thus engage the user on a categorically different level compared to interfacing only via written text and the 2D display of an animated face.

2.1 Problems of the physical world

With the benefits of the step into the physical world, however, come the difficulties of the physical world. Not only becomes perfect virtual perception noisy real world sensing, precise and almost delay-free visual animation imprecise and execution time-adherent physical activation, but also the stakes are set higher to achieve the ultimate goal of creating a believable interactive agent. The virtual world is (at least currently) much sparser than the physical world and thus offers substantially less cues to the observer. Less cues mean less opportunities to destroy the user's perception of agency which is fragile no matter how sophisticated the underlying control algorithms might be given the current state of art of artificial intelligence. In other words, compared to the virtual-only agent, many more aspects of the robotic agent must be modeled correctly, because failure to do so would immediately

expose the 'dumb' nature of the agent. This might not constitute a problem in some of the applications of human-robot collaboration we discussed above since the human co-worker might easily accommodate to shortcomings and peculiarities of the machine colleague but it can be assumed that in many other contexts the tender fabric of interactions will be torn apart, in particular, if the interactions are more complex. Statements in this regard are currently marred by their speculative nature as the appropriate research using psychological experiments has not been done yet. This is equally due to the lack of sufficiently advanced and interactive robots as to the difficulties to even simulate such a robot and systematically vary experiment conditions in so-called Wizard-of-Oz experiments where unknown to the participants a human operator steers the robot.

In our case of a robotic conversational agent, the overall goal of the art project was at stake: the ability to engage in a conversation, to take turns in a dialogue, to use language and speech more or less correctly, requires as a prerequisite an intentional agent. Thus, if the robot's actions had betrayed the goal of evoking and maintaining the impression of intentionality and agency, it would have compromised the agent as a whole: either by unmasking the integrated chatbot as a 'shallow' algorithm operating on a limited database with no deeper understanding of the content of the dialogue or by destroying the perception of embodiment by introducing a rift between the 'clever' language module and the failing robot.

2.2 Convincing behaviour

The cardinal problem encountered is the requirement to respond to a changing stimulus-rich environment with reasonable and appropriate behaviour as judged by the human observer. Overcoming this problem is not possible, we propose, without integration of the plenitude of incoming sensory information as far as possible and selection of the most relevant parts taking into account that (for our purposes) the sensory information is not a sufficiently complete description of the physical environment. Therefore, as a first step after low-level sensory processing, an attention mechanism is necessitated that prioritises information relevant to the current task of the agent over less important incoming data. An attention model not only takes care of the selection process, it also implicitly solves the problem of a vastly incomplete representation of the environment. For any control system that receives the output of the attention model, it is per se evident that it receives only a fragment of the available information and that, should this information not be sufficient for the current task, further data need to be actively acquired. In a second step then, the selected stimuli have to be responded to with appropriate behavior, which means in most cases with motor action though at other times only the settings of internal state variables of the system might be changed (e.g. an attention threshold).

There is another important issue here: when it comes to the movements of the robot not only the 'what' but also the 'how' gains significance. 'Natural' movements, i.e. movements that resemble biological movements, contribute crucially to the overall impression of agency as the Articulated Head has a realistic human face. Robot motion perceived as 'mechanical' or 'machine-like' would abet the separation of the robot and the virtual agent displayed on the LCD monitor, and thus create the impression of a humanoid agent being trapped in the machine. Again, if we allow a little bit of speculation, it can be hypothesised that robot arms engaging in joint action with humans will need to generate biological motions in order to make predictions of future actions of the robot arm easier and more intuitive for the collaborating human.

Joint	1	2	3	4	5	6
Motion range (deg)	340	200	388	380	240	720
Motion speed (deg/s)	350	350	400	450	450	720

Table 1. Robot joint motion range and speeds.

3. The robot arm

The robot arm employed is a Fanuc LR Mate 200iC used typically in industrial applications. It has six degrees of freedom. Table 1 shows the speed and the motion range of each individual joint of the robot arm with joint 1 being the closest to the mounting base. The robot is mounted on a custom made four-legged heavy steel structure which does not require fixing it to the floor for stable operation. The robot's work envelop is protected from inadvertent entry by human users through a series of glass structures and interlocks. The robot is controlled through an external control box using a proprietary handling application program language. In standard usage the robot is pre-programmed with the necessary movement instructions using a 'teaching pendant' with target points entered 'online' during the teaching phase prior to commissioning of the robot. However, in order to accommodate realtime interactive behavior desired by the Articulated Head, the robot interface was customised such that target points (i.e. motor goals, see section 7) could be created during the execution phase. This also meant that an additional layer of safety checks were needed to prevent the robot from trying to reach unreachable locations resulting in collisions and/or singularity conditions.

4. Inter-Component Communication and Sensing

4.1 Software Architecture

The communication framework (Herath et al., 2010) for our system combines approaches from open agent-oriented systems previously used for multimodal dialogue systems (e.g. Herzog & Reithinger, 2006) and frameworks for high-performance robotic platforms (e.g. Brooks et al., 2005; Gerkey et al., 2003). The driving motivation is to enable easy integration of components with different capabilities, written in different programming languages and potentially running on different platforms (including distributed platforms). A specific requirement for our application is realtime performance under massive data processing over streaming audio and video; this ruled out the existing multimodal dialogue platforms, and also led us to eschew standards-based APIs which incur overheads on message-passing to components. In common with other dialogue platforms, we use an event-driven framework, which has a number of desirable properties, such as: naturally modelling the non-linear nature of human interaction; providing the flexibility required for easy integration of components into a distributed architecture; dynamically prioritising software components and event types; and optimising the system via inter-component configuration commands for particular interaction states.

4.2 Sensors

We have adopted two commercially available camera systems for tracking people in 3D and faces in close proximity. A stereo camera mounted rigidly high on a wall opposite the Articulated Head looking downwards into the interaction space of the robot provides information about human movement. The commercial people tracking algorithm is based

on an assumed depth profile of an average human and uses disparity images produced by a calibrated camera pair. It provides the localisation and height information of all people within the camera's field of view to the robot. The tracking system is capable of tracking multiple persons with considerable tolerance to occlusion and occasional disappearance from the field of view.

A monocular camera mounted above the top edge of the LCD screen provides fine grain information about humans directly interacting with the robot. Data from this camera feeds on to a face tracking algorithm that is capable of detecting and tracking a single face in the camera's field of view. The used algorithm has a high degree of accuracy withstanding considerable occlusion, scale variance and deformations.

Stereo microphones mounted on the back panel of the robot enclosure coupled to an auditory localiser provides accurate information of the instantaneous locations (azimuth) of a moving interlocutor in a noisy and reverberant environment. Localisation is limited to the half sphere in front of the robot and provides azimuth angle from about -90° to $+90^\circ$. The localisation is based on Faller & Merimaa (2004) which has been modified and adapted to the Articulated Head setup.

In addition to above components, various ancillary components such as proximity detectors, keyboard input device, gesture recognition system, text-to-speech system, dialogue management system, monitoring and a data logging systems are also implemented to support the various interactive aspects of the Articulated Head. Figure 3 shows the overall component topology.

5. Attention model

Human attention is a heavily researched area with thousands of scholarly articles. In general, attention is investigated in controlled psychological experiments focusing on specific aspects of the overall phenomenon of attention, say, visual attention activated by certain types of motion perceived in peripheral vision. A substantial amount of knowledge has been accumulated, though sometimes disparate or conflicting. One of the most important findings for attention systems in machines (Shic & Scassellati, 2007) is that attention is driven by two sources, saliency in the perceptual input (bottom-up or exogenous attention) and task-dependent attention direction (top-down or endogenous attention). The former is comparatively easier to handle and evaluate (e.g. data from human participants can be acquired using eye-tracking technology). Top-down attention mechanisms, on the other hand, pose severe difficulties as they typically involve high-level world knowledge and understanding. Unfortunately, however, top-down mechanisms appear to be more decisive: Even for a barn owl only 20% of attentional gaze control could be explained by low-level visual saliency (Ohayon et al., 2008).

Compared to human attention, modelling attention in artificial agents is less studied. Attention models have been primarily investigated and applied in virtual environments (Kim et al., 2005), avoiding the largely unsolved problem of real world object recognition. The identity of objects placed in a virtual environment can be made known directly to the attention model of the agent; an option that is clearly not available when dealing with a robot and real world sensing. In addition, the sensory input in real world sensing is always affected by considerable noise. The majority of the attention models for virtual agents is biologically inspired and thus complex (e.g. Bosse et al., 2006; Itti et al., 1998; Peters & Itti, 2006; Sun et al., 2008), though others amount to not much more than a fixed selection process of an input source based on the value of a single (or a few) parameter.

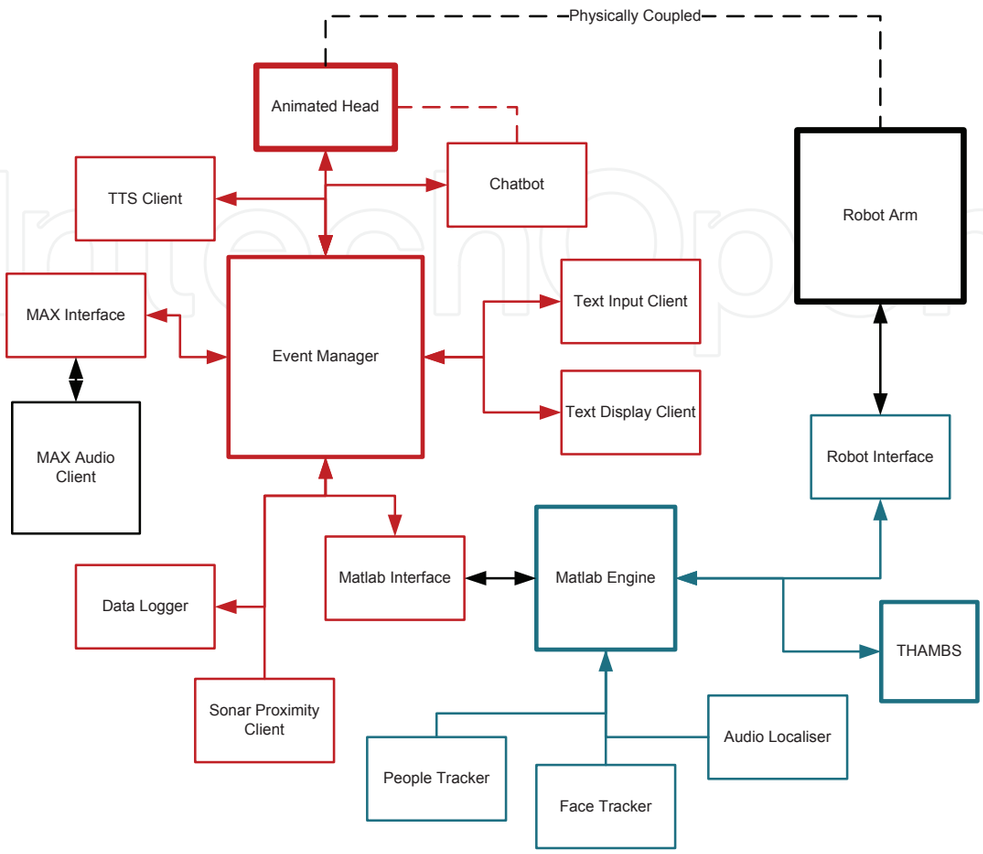


Fig. 3. Component topology. 'THAMBS' in the rightmost box stands for the Thinking Head Attention Model and Behavioural System and is described in section 5 to 8.

5.1 Attention models in robotics

A few attempts have been made to develop attention models for robots. One of the first implementations, named FeatureGate, used an artificial neural network that operated on 2D feature maps (Driscoll et al., 1998), specifically - in the tests presented in the paper - feature maps derived from synthetic images. In its handling of how the features were weighted, it allowed changes depending on the task, that is, top-down attention was partially established. However, despite its sophisticated algorithms, FeatureGate corresponds more to a target-detection system than an attention system as it does almost nothing other than find a given target among distractors in an efficient manner. This is in line with many of the experiments studying visual attention in humans, but these experiments use a simplified controlled experiment set-up to isolate aspects of the complex human attention system; they do not indicate that the human attention can be reduced to an efficient search method (e.g. Cavanagh, 2004). We would argue that when it comes to work with a robot, attention truly starts when there are several potential targets and the system has to make a choice: discard (temporarily) all but one target (the most relevant one given the current task) and focus on it.

Only bottom-up attention mechanisms were considered in the attention model developed in Metta (2001). The study focused on log-polar vision which simulates the distribution of the photoreceptors in the primate eye. They showed that this type of space variant vision is well suitable for implementing an attention system and controlling robot movements through it. It addresses implicitly the reduction of previous attention systems to target detection systems by having different sensory resolutions in the periphery and the foveal area. Thus, attention coincides with the fixation point, but events registered in the periphery could still attract attention and command fixation.

Like the model of Driscoll and colleagues, the visual attention system of Breazeal & Scassellati (1999) used with the robot Kismet was based on the guided search model of Cave & Wolfe (1990) and Wolfe (1994), but it went beyond it. The attention system did not only combine several different feature maps, but also modeled the influence of habituation effects and integrated the impact of the robot's motivational state on the generated attention activation map. In this way, their attention system became context-dependent and Kismet's behaviour emerged from the interaction of its own state and the state of the environment. For instance, the attentional gain for faces was increased during Kismet's *seek people* behaviour and decreased during its *avoid people* behaviour. Kopp & Gärdenfors (2001) postulated the imperative of an attention system for perceived intentionality of a robotic agent, but, unfortunately, they did not implement one. Their robot arm equipped with two cameras, one for peripheral vision (above arm) and one for central/focal vision (at arm), would have been, as they indicated, a very suitable platform for it.

A multimodal attention system to guide an interactive robot was proposed in Déniz et al. (2003). The researchers used feature and saliency maps to model bottom-up attention combining visual and acoustic features, but did not include top-down processes. They also did not treat visual and acoustical events equally: acoustic events could not change the focus of attention, they only reinforced the visual event closest to the acoustic event.

Attention models based on salience maps (the majority of those mentioned above) can be computationally very costly, particularly if an increasing number of features and larger feature and salience maps are used. Ude et al. (2005) demonstrated that with proper parallel processing in a distributed implementation, sufficient speeds were achieved to steer the visual system of the humanoid robot they used in realtime. The model of Ude et al. (2005) was further developed in (Morén et al., 2008) by strengthening the top-down aspects and exploring a new way of integrating bottom-up and top-down mechanisms. The authors combined the use of saliency maps from Itti and Koch's (Itti et al., 1998) model with a more flexible version of the feature-specific top-down mechanism of Cave's FeatureGate (Cave, 1999).

In this vein it appears as if attention models in robots have been recently recognised as a way to tackle problems with visual segmentation. As we mentioned earlier, this view seems at times to be more inspired by psychological experiments investigating visual attention than biological attention itself; they seem to model aspects of those experiments. As a consequence, robotic attention does not only fail to model the complexity of human attention - something which is expected and generally unavoidable given the current state of technology - but also reduces attention to an auxiliary function of the robot's perceptual system while it should be, if anything, its 'guide'. Nevertheless useful results can be obtained. Yu et al. (2007), for instance, devised an attention-based method to segment specified object contours from the image motion produced by the egomotion of a mobile robot. They employed a pre-attentive state for contour segmentation and competing motion-based bottom-up and contour-based top-down

salience maps. Using Bayesian inference top down saliency biased the final probabilistic attention distribution toward the task-dependent object contour.
Developing and applying attention models is usually motivated with the proposed requirement that robots interacting with humans should possess an attention system similar to that of humans. More specifically, *joint* attention is argued to be a conditio sine qua non for cooperative human-robot action, machines learning from human instructors, 'theory of mind' in robots (the ability to predict what another can and cannot perceive), and similar high-level cognitive social capabilities. However, joint attention of robot and human is trivially not possible if the robot does not have the capacity of attention in the first place (at least in the form of being able to select specific elements of the input over others). In the literature reviewed above attention systems sometimes seem more to be a means to an auxiliary end than being an integral and essential part of the robots behavioural system. A different and more immediate motivation is brought forward in Bachiller et al. (2008) in their 'attention-based control model'. The authors view the attention system as an essential mediator between visual perception and action control that is needed to handle two important tasks: to select perceptual information relevant for action execution and to limit potential actions based on the perceived situational context. In the context of autonomous navigation of robots they employ bottom-up and top-down attention processes and also model overt and covert attention. Covert attention refers here to regions of interest that are pre-activated within the attention system through target selection, but are currently not the focus of attention (overt attention).
Finally, a method to switch autonomously between bottom-up and top-down attention in a mobile robot was introduced by Xu et al. (2010). The different attention modes are activated dependent on the state the robot is in (exploring, searching, or operating) which links attention back to behaviour - something that in our view is essential for attention: attention cannot be seen as a passive input information selection mechanism since it is tied to action and also actively changes what is perceived. The benefits of the latter was demonstrated in Xu et al. (2010) through steering the active stereo camera of the robot they used towards target area identified by the bottom attention system as relevant for the task and then apply the top-down attention to keep the target in the focus of attention.

5.2 The attention model of the Articulated Head

In the Articulated Head, the attention model is part of the Thinking Head Attention and Behavioural System (THAMBS) that manages all high-level aspects of the interaction including the generation of response behaviour (see next section) except for conversational matters that are taken care of by the chatbot. THAMBS goes beyond straightforward action selection insofar as it is also concerned with determining the specific characteristics of the motor behaviour associated with the response (see section 7) and in that behaviours can interact with each other and can change the way the sensory input is processed. It consists of four modular subsystems: (1) a perception system, (2) an attention system, (3) a central control system, and (4) a motor system. Figure 4 shows a diagram of THAMBS, with its subsystems and flow of information. THAMBS is currently implemented in Matlab (The MathWorks, Inc) and following object-oriented programming principles, its subsystems are represented as classes to ensure their strict modularity.
Despite the array of sensing devices, the robot's sensing of the world is relatively sparse since the sensing devices and their software are specialised on particular tasks (e.g. people tracking, face detection). It is multimodal, however, and complex enough to allow sophisticated interactions with human users. Nevertheless, the difference in the input compared to almost

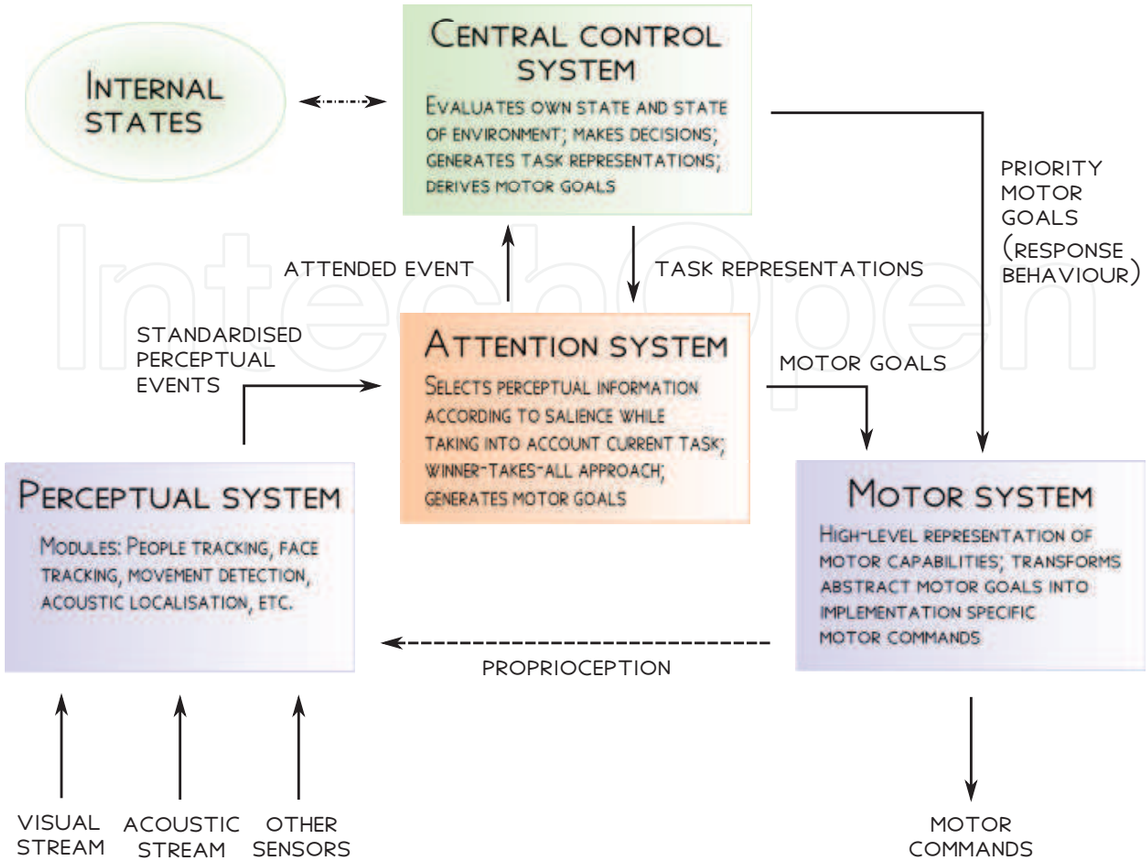


Fig. 4. Schematic of THAMBS and its subsystems.

all other attention models has implications on architecture and functionality of the attention system within THAMBS. First an upstream perception system transforms the information from the sensing modules into a standardised perception event making it possible to process very different events (e.g. a person being detected within the visual field of the Articulated Head as well as a character string being sent from the keyboard) with respect to their attentional importance not their detail characteristics. An attentional weight is assigned to the incoming perceptual event computed using a base weight assigned as a parameter value to the *type* of the perceptual event (e.g. acoustic localisation, people tracking) and an attention weight factor derived from the specific event *instance*, usually confidence values (the default value is 1):

$$w_p(i) = b_p(i)\, w_{base} \tag{1}$$

Note that both values might be changed during run time according to changes of the active task, for instance the base weight of face tracking might be increased to favour face-to-face interactions with a single person over 'distracting' other people in the area covered by the stereo camera. The resulting attention weight is checked against a threshold dependent again on the type of perceptional event. If the event passes, an attention focus is created (covered

attention) which is characterised by three properties: its weight, its decay function, and the spatial location in the real world it is referring to.
The weight is the attentional weight described above, however, for an already existing focus it is modified based on the duration of its existence.

$$w_p(i,t) = w_p(i)\ k(t) \tag{2}$$

where $k(t)$ is a decay function assigned to the attention focus that decides about its lifetime and its impact over time. It ensures that the attention focus outlives the potentially very short instance of the perceptual event that created it (e.g. a loud startling impulse-like noise) but at the same time that its strength is fading even if registration of the perceptual event is sustained (habituation). We found a generalisation of the simple exponential functions, the Kohlrausch function, preferable to the simple exponential functions employed in other attention models. The Kohlrausch function is often called a 'stretched exponential' and is known to be able to describe a wide range of physical and biological phenomena (Anderssen et al., 2004). It is given by:

$$k_{\tau,\beta}(t) = e^{-(\frac{\tau}{\beta})^{\beta}} \tag{3}$$

It is the additional parameter β that stretches or compresses the function. Thus with an appropriate setting a plateau at around zero is formed that guarantees in our implementation a high activation for a certain period immediately after the attention focus has been established ensuring that the focus can 'fend off' lower weight foci for some time. The decay function parameters are initialised dependent on the type of perceptual event, but again, they are modified dynamically during run time (in fact, the entire function can be replaced with a different one if e.g. a discontinuous function is needed. However, this is currently not used).
The last and most important defining property of an attention focus is the segment of 3D space it is referring to: the location of the event that attracted attention. Thus, the attention foci are spatially organised (compare space versus object-based attention in models of human attention; review in Heinke & Humphreys, 2004). This plays a decisive role in the identification of a new perceptual event as identical - per definition - to one of the already existing attention foci. Locations in spherical coordinates of the new event and all old foci are compared whereby underspecification always produces a positive value. If an incoming event is considered to be identical to one of the old attention foci, the old focus is kept. Its weight, however, is updated by combining of the new and old weights in supra-additive manner:

$$w_c = w_{old} + w_{new}\ a_p \qquad \text{with} \quad 0 \leq a_p \leq 1 \tag{4}$$

where a_p is a parameter specific to the perceptual event type of the new event. The decay function of the focus is not reset, which causes a slow but steady decline of the weight values even if new events are constantly reinforcing an old attention focus, for instance, a person standing still within the visual field of the Articulated Head. The procedure has a similar effect as the habituation modeled in Breazeal & Scassellati (1999).
A perceptual event might have several different features depending on its type. There are always the obtained sensory data values themselves, typically some form of tracking data (though in case of the keyboard 'sense', it is only a binary on/off signal and a character string) but in addition there might be velocities and, potentially, accelerations or other properties computed over the input values such as statistical and spectral moments or energy measures. Each feature on its own is able to invoke an attention focus: If one feature fails to create an attention focus because it can not pass the threshold, another feature might do so. For instance,

an avoidance behaviour might have blocked tracked people to become an attention focus, but very fast movements of a person might nevertheless 'break through' via the velocity feature. After all attention foci are generated and their weights computed, one of them is chosen as the single event that is attended by the system using a winner-takes-all-strategy based on the highest weight. Recently, we added an alternative, a persistence strategy. After a perceptual event providing an identification marker (currently only people tracking) has become the attended event through the default strategy, the attention system locks on the event based on its ID for a limited time - independent of its decaying weight - unless there is a very powerful distractor. The trigger for the persistence strategy is random at the time, but it was devised to be replaced with a trigger based on familiarity once face recognition is integrated in THAMBS. The attended event is passed on to the central control system (see next section) together with the information whether or not it is considered new by the attention system. To model pursuit movements based on a close attentional link between perception and action (Schneider & Deubel, 2002), the attention system is able to send a specific motor command, named `look_there`, directly to the motor system. It steers the robot arm to orient the normal to the monitor display plane (and with this the optical axis of the monovision camera) toward the spatial location of the attended event. This serves a two-fold purpose: to create the impression as if the virtual face displayed on the monitor is looking at the location of the event which attracted its attention and to provide the Articulated Head with more information about the source of the event via the monovision camera (see Figure 5).

(a) Idle (b) Idle (c) Focussing

Fig. 5. The Articulated Head being idle (a,b) and focussing on an interlocuter (c).

6. Behavioural system

The response behaviour of the Articulated Head is generated by the central control subsystem of THAMBS (except for verbal interactions). It is the highest-level processing stage for information about the environment that arrives from the sensors after being evaluated by the attention system. This information itself, however, is not independent from the behaviour (perception-action link) as the behaviour affects the sensing information either directly (e.g. the position and orientation of the monovision camera) or indirectly as the behaviour might cause a change in task priorities which in turn might trigger a modification in the attentional weights assigned to perceptual event types or single attention foci. The central control system is essentially still a stimulus-response system based on a set on conditional rules, but it is non-trivial since the rules are modified during run time and are at some points subject to probabilistic evaluation.

6.1 Behaviour triggers

In THAMBS the conditional rules are called *behaviour triggers* and realised as small decision trees. In most current triggers, however, only one branch leads to the activation of a behaviour while the remaining ones cause a termination of the trigger evaluation. This is bound to change with more complex behaviour options being implemented at future development stages. The trigger evaluation is implemented to be able to handle trees of arbitrary complexity fast and efficiently while requiring only a few lines of code in Matlab. The basic idea is to collect the test results as '0' or '1' characters in a single string while moving down the tree following the active branch and then treat the resulting string as representing a binary number and convert this number into an index into the possible actions associated with the terminal nodes. In pseudo-code:

```
(1) Collect all tests associated with the nodes of the decision tree
in a one-dimensional array of expressions (named 'conditions' here)
that evaluate to a Boolean value. Move from top to bottom and left to
right.
(2) Collect from left to right all possible actions associated with
the terminal nodes in a one-dimensional array of function handles
(named 'actions' here).
(3) Initialise an indicator variable 'indTest' with value 1,
another indicator variable 'indAction' with 0, and an empty string
array 'collectedTests'.
Note: It is assumed that array indexing starts with 1 not 0.

%%% loop through all (relevant) tests of the decision tree.
WHILE indTest <= SIZE(conditions)

  %%% evaluate the indicated condition
  isTrue = EVALUATE(conditions[indTest])

  %%% if condition evaluates to true, add a '1' else a '0' to the string
  %%% array that collects the test results
  IF isTrue
    add '1' to collectedTests
  ELSE
    add '0' to collectedTests
    %%% stop further testing in case of a branch type tree
    %%% (only one branch of the tree has a behaviour assigned)
    IF tree is of type branch
      BREAK
    ENDIF
  ENDIF

  %%% split according to type of tree: determine the index for the
  %%% next test
  IF tree is of type branch
     indTest = indTest + 1
  ELSE
     %%% treat the collected test string buffer as a binary number and
     %%% use it to update the index. In this way tests that belong to
     %%% branches, which have been already discarded, will be ignored
     indTest = 2 ^ (SIZE(collectedTests)) + BINARY_TO_DECIMAL(collectedTests)
  ENDIF
ENDWHILE
```

```
%%% treat the collected test string buffer as a binary number
%%% and convert it to a decimal number: the result is an index in the
%%% possible actions arranged according to the last level of the tree
indAction = BINARY_TO_DECIMAL(collectedTests) + 1;

%%% for type 'branch' only one branch has an action assigned: do we have
%%% it? If yes, set indAction to 1
IF tree is of type branch
  IF indAction == 2^SIZE(conditions)
    indAction = 1
  ELSE
    indAction = 0
  ENDIF
ENDIF

%%% Call the appropriate action function
IF indAction > 0
  EXECUTE(actions[indAction])
ENDIF
```

Behaviour triggers have a priority value assigned to them which decides about the order of their evaluation. A trigger with a higher priority is evaluated after a trigger with a lower priority since the associated behaviours of both might modify state variables of THAMBS and the changes made by the behaviour with the higher priority trigger should take precedence, that is, these changes should be the ones that persist and should not be overwritten. Typically behaviours specify a motor goal to be achieved. Motor goals are abstract representations of motor actions to be executed by the robot arm or the virtual head displayed on the monitor. They are context independent and thus no sensory information is required at this stage, though several attributes control their processing by the motor system later on. Other behaviours only modify values of state variables and with this cause a change in how future sensory information is processed or in how other motor goals are executed. The behaviours themselves are implemented as independent routines and their function handles are passed to the trigger evaluation routine.

6.2 Behaviour disposition

Two other important aspects of the central control system besides the generation of response behaviour need to be mentioned. First there is a set of subroutines that model endogenous processes. These are changes in THAMBS' state variables that are not activated - directly or indirectly - by stimuli from the environment. An example would be the spontaneous probability-driven awakening that happens sooner or later if the Articulated Head has fallen asleep (due to lack of stimuli in the environment; see section 9 for an overview of the behaviors of the Articulated Head). It contrasts with the awakening activated by a loud sound event via an ordinary behaviour trigger (Kroos et al., 2010). The endogenous processes would be more accurately assigned to a system other than the central control system as they emulate low-level functions of the mammalian brain located, for instance, in the brain stem. Future versions of THAMBS will parcel out these processes and subordinate them to a new subsystem.
Secondly, there is a preparatory phase. THAMBS currently employs a master execution loop running usually at 10 Hz through all the necessary tasks of its systems, starting with the endogenous processes, then handling perception, continuing with attention and so on.

However, we mentioned in section 5 that attention has a strong top-down component. This is specifically accounted for in the preparatory routine which is executed before the perception system becomes active in each evaluation cycle of the master loop. The routine can change thresholds of perception and attention, and in this way it can steer perception and attention toward stimuli relevant for its current task and its current inner state (active perception and active attention). Moreover, it is able to insert new behaviour triggers in the set of active behaviour triggers. For instance, the behaviour trigger `attend_close` activates a behaviour with the same name if a sizable number of people are in the visual field of the Articulated Head. The `attend_close` behaviour changes the weight of attention foci that are based on people-tracking to favour people closer to the Articulated Head over people further away. The trigger has limited lifetime and is currently inserted randomly from time to time. In future versions this will be replaced by an insertion based on the values of other state variables, e.g. the variable simulating anxiety. Note that the insertion of an behaviour trigger is not equivalent with activation of the associated behaviour. Indeed, taking the example above, the `attend_close` behaviour might never be activated during the lifetime of the trigger if there are no or only few people around. An Articulated Head made 'anxious' through the detection of a reduction in computational resources might insert the behaviour trigger fearing a crowd of people and dealing with this 'threatening' situation in advance.
The distinction between preemptive behavior disposition and actual response triggers is important because it constitutes an essential element in the differentiation of a simple context-independent stimulus-response system with the classical strict division of input and output from an adaptive system where the interaction with the environment is always bi-directional. Note also that the preparatory phase de-facto models expectations of the system about the future states of its environment and that contrary to the claims in Kopp & Gärdenfors (2001), this does not necessarily require full internal representations of the environment.

7. Motion generation

The motor subsystem of THAMBS is responsible for converting the abstract motor goals transmitted both from the attention system and the central control system into concrete motor primitives. At first, the motor system determines which one of the two motor goals - if both are in fact passed on - will be realised. In almost all cases the 'deliberate' action of the central control system takes precedence over the pursuit goal from the attention system. Only in the case of an event that attracts exceptional strong attention the priority is reversed. In humans, this could be compared with involuntary head and eye movements toward the source of a startling noise or toward substantial movement registered in peripheral vision. A motor goal that cannot currently be executed might be stored for later execution depending on a specific storage attribute that is part of the motor goal definition. For pursuit goals originating from the attention system the attribute is most of the time set to disallow storage as it makes only limited sense to move later toward a then outdated attention focus. On completion of the goal competition evaluation, the motor systems checks whether the robot is still in the process of executing motor commands from a previous motor goal and whether this can be interrupted. Each motor goal has an `InterruptStrength` and an `InterruptResistStrength` attribute and only if the value of the `InterruptStrength` attribute of the current motor goal is higher than the `InterruptResistStrength` of the ongoing motor goal, the latter can be terminated and the new motor goal realised. Again, if the motor goal cannot currently be executed it might be stored for later execution.

Motion generation in robot arms might be considered as a solved problem (short of a few problems due to singularities maybe) and as far as trajectory generation is concerned we would agree. The situation, however, changes quickly if requirements on the meta level of the motion beyond desired basic trajectory properties (e.g. achieving target position with the end effector or minimal jerk criteria) are imposed. In particular in our case, as mentioned in section 2.2, the requirement of the movements to resemble biological motion. Since there exists no biological model for joint system such as the Fanuc robot arm, an exploratory trial-and-error-based approach had to be followed. At this point a crucial problem was encountered: if the overall movement of the robot arm was repeated over and over again, the repetitive character would be quickly recognised by human users and perceived as 'machine-like' even if it would be indistinguishable from biological motion otherwise. Humans vary constantly albeit slightly when performing a repetitive or cyclical movement; they do not duplicate a movement cycle exactly even in highly practised tasks like walking, clapping or drumming (Riley & Turvey, 2002). In addition, the overall appearance of the Articulated Head does not and cannot deny its machine origin and is likely to bias peoples' expectations further. Making matters worse, the rhythmical tasks mentioned above still show a limited variance compared to the rich inventory of movement variation used in everyday idle behaviour or interactions with other people - the latter includes adaptation (entrainment) phenomena such as the adjustment of one's posture, gesture and speaking style to the interlocutor (e.g. Lakin et al., 2003; Pickering & Garrod, 2004) even if it is a robot (Breazeal, 2002). These situations constitute the task space of the Articulated Head while specialised repeated tasks are virtually non-existent in its role as a conversational sociable robot: one more time the primary difference between the usual application of a robot arm and the Articulated Head is encountered. Arguably, any perceivable movement repetition will diminish the impression of agency the robot is able to evoke as much as non-biological movements if not more.

To avoid repetitiveness we generated the joint angles for a subsets of joints from probability density function - most of the times normal distributions centred on the current or the target value - and used the remaining joints and the inherent redundancy of the six degrees of freedom robot arm to achieve the target configuration of the head (the monitor). Achieving a fixed motor goal with varying but compensating contributions of the participating effectors is known in biological motion research as *motor equivalence* (Bernstein, 1967; Gielen et al., 1995). The procedure we used not only resulted in movements which never exactly repeat but also increased the perceived fluency of the robot motion.

Idle movements, small random movements when there is no environmental stimulus to attract attention, are a special case. No constraint originating from a target configuration can be applied in the generation of these movements. However, completely random movements were considered to look awkward by the first author after testing them in the early programming stages. One might speculate that because true randomness is something that never occurs in biological motion, we consider it unnatural. As a remedial, we drew our joint angle values from a logarithmic normal (log normal) distribution with its mean at the current value of the joint. As can be seen in Figure 6, this biases the angle selection toward smaller values than the current one (due to a cut-off at larger values forced by the limited motion range of the joint; larger values are mapped to zero), but in general keeps it relatively close to the current value. At the same time in rare cases large movements in the opposite direction are possible.

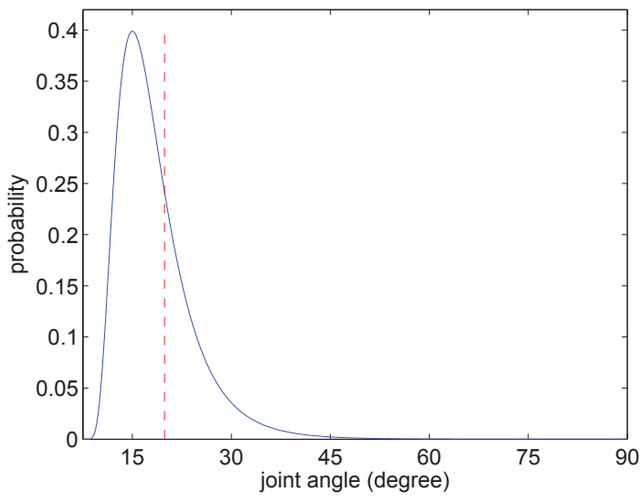


Fig. 6. Log normal probability distribution from which the new joint angle value is drawn. The parameters of the distribution are chosen so that the mean coincides with the current angle value of the robot joint. In this example it is at 24.7 degree indicated in the figure as dotted line and the cut-off is set to 90 degree.

The generation of the motor primitives realising an abstract motor goal is handled by specialised execution routines. The handles to these functions are stored as motor goal attributes and can be exchanged during runtime. The subroutines request sensory information if required such as the location of a person to be 'looked at' and transduce the motor goal in the case of the robot arm into target angle specifications for the six joints, and in case of the virtual head into high-level graphic commands controlling the face and eye motion of the avatar.
The joint angle values determined in this way are sent to the robot arm after they have passed safety checks preventing movements that could destroy the monitor by slamming it into one of the robot arm's limbs.

8. State variables and initial parameters

We described THAMBS from a procedural point of view which we deemed more appropriate with respect to the topic of evoking agency and more informative in general. However, this does not mean that there is not a host of state variables that provide the structure of THAMBS beyond the subsystems described in the previous section. In particular, the central control system has a rich inventory of them. They are organised roughly according to the time scale they operate on and their resemblance to human bodily and mental states. There are (admittedly badly named) 'somatic' states which constitute the fastest changing level, then 'emotional' states on the middle level and 'mood' states on the long term level. Except for the somatic states such as alertness and boredom those states are very sparsely used for the time being, but will play a greater role in further developments of THAMBS.
Although the behaviour of the Articulated Head emerges from the interplay of environmental stimuli, its own actions, and some pre-determined behaviour patterns (the behaviour triggers described in section 6.1), a host of initial parameter settings in THAMBS influences the overall behaviour of the Articulated Head. In fact, very often changing individual parameter settings creates patterns of behaviour that were described by exhibition visitors in terms of different

personalities or sometimes mental disorders. To investigate this further, however, a less heuristically driven approach is needed for modelling attention and behaviour control and rigorous psychological experiments. At the time of the writing both are underway.

9. Overview of most common behaviour patterns

If there is no environmental stimulus strong enough to attract the attention of THAMBS, the Articulated Head performs idle movements from time to time and the value of its boredom state variable increases. If it exceeds a threshold, the Articulated explores the environment with random scanning movements. While there is no input reaching the attention system, the value of the alertness state variable decreases slowly such that after prolonged time the Articulated Head falls asleep. In sleep, all visual senses are switched off and the threshold for an auditory event to become an attention focus is increased. The robot goes into a curled-up position (as far as this is possible with the monitor as its end effector). During sleep the probability of spontaneous awakening is very slowly increased starting from zero. If no acoustic event awakens the Articulated Head it wakes up spontaneously nevertheless sooner or later. If its attention system is not already directing it to a new attention focus, it performs two or three simulated stretching movements.

If there is only a single person in the visual field of the Articulated Head, it focuses in most instances on this person and pursues his or her movements. There might be, however, distractions from acoustic events if they are very clearly localised. If the person is standing still, the related attention focus gains for a short time a very high attentional weight, but if nothing else contributes, the weight fades, making it likely that the Articulated Head diverts its attention. Alternatively, the face detection software might register a face as the monovision camera is now pointing toward the head of the person and the person is not moving anymore. This would lead to a strong reinforcement of the attention focus and in addition the Articulated Head might either speak to the person (phrases like 'I am looking at you!', 'Did we meet before?', 'Are you happy?' or 'How does it look from your side?') or mimic the head posture. The latter concerns only rotations around the axis that is perpendicular to the monitor display plane in order to be able to maintain eye contact during mimicry.

If a visitor approaches the information kiosk (see Figure 7) containing the keyboard, the proximity sensor integrated into the information kiosk registers his or her presence. The Articulated Head turns toward the kiosk with a high probability because the proximity sensor creates an attention focus with a high weight. If the visitor loses the attention of THAMBS again due to inactivity or sustained typing without submitting the text, the Articulated Head would still return to the kiosk immediately before speaking the answer generated by the chatbot.

If there are several people in the vicinity of the Articulated Head, its behaviour becomes difficult to describe in general terms. It now depends on many factors which in turn depend on the behaviour of the people surrounding the installation. THAMBS will switch its attention from person to person depending on their movements, whether they speak or remain silent, how far they are from the enclosure, whether it can detect a face and so on. It might pick a person out of the crowd and follow him or her for a certain time interval, but this is not guaranteed when a visitor tries to actively invoke pursuit by waving his or her hands.

Fig. 7. The information kiosk with the keyboard for language-based interactions with the Articulated Head.

10. Validation

The Articulated Head is a work of art, it is an interactive robotic installation. It was designed to be engaging, to draw humans it encounters into an interaction with it, first through its motor behaviour, then by being able to have a reasonably coherent conversation with the interlocutor. Because of the shortcomings of current automatic speech recognition systems (low recognition rates in unconstrained topic domains, noisy backgrounds, with multiple speakers) a computer keyboard is still used for the language input to the machine but the Articulated Head answers acoustically with its own characteristic voice using speech synthesis. It can be very entertaining but entertainment is not its primary purpose but a consequence from its designation as a sociable interactive robot. In terms of measurable goals, interactivity and social engagement are difficult to measure, in particular in the unconstrained environment of a public exhibition.

So far the Articulated Head has been presented to the public at two exhibitions as part of arts and science conferences (Stelarc et al., 2010a;b) and hundred of interactions between the robotic agent and members of the audience have been recorded. At the time of the writing, a one year long exhibition in the Powerhouse Museum, Sydney, Australia, as part of the Engineering Excellence exhibition jointly organised by the Powerhouse Museum, Sydney, and the New South Wales section of Engineers Australia has just started (Stelarc et al., 2011). A custom-built glass enclosure was designed and built by museum staff (see Figure 8) and a lab area immediately behind the Articulated Head installed allowing research evaluating the interaction between the robot and members of the public over the time course of a full year.

Fig. 8. The triangular-shaped exhibition space in the Powerhouse Museum, Sydney.

This kind of systematic evaluation is in its earliest stages, preliminary observations point toward a rich inventory of interactive behaviour emerging from the dynamic interplay of the robot system and the users. The robot's situational awareness of the users' movements in space and its detection of face-to-face situations, its attention switching from one user and one sensory systems to the next according to task priorities that is visible in its expressive motor behaviour, all this entices changes in the users' behaviour which, of course, modify again the robots' behaviour. At several occasions, for instance, children played games similar to hide-and-seek with the robot. These games evolved spontaneously despite that they were never considered as an aim in the design of the system and nothing was directly implemented to support them.

11. Conclusion and outlook

Industrial robot arms are known for their precision and reliability in continuously repeating a pre-programmed manufacturing task using very limited sensory input, not for their ability to emulate the sensorimotor behaviour of living beings. In this chapter we have described our research and implementation work of transforming a Fanuc LR Mate 200iC robot arm with an LCD monitor as its end effector into a believable interactive agent within the context of a work of art, creating the Articulated Head. The requirements of interactivity and perceived agency imposed challenges with regard to the reliability of the sensing devices and software, selection and integration of the sensing information, realtime control of the robot arm and motion generation. Our approach was able to overcome some but certainly not all of these challenges. The corner stones of the research and development presented here are:

1. A flexible process communication system tying sensing devices, robot arm, software controlling the virtual avatar, and the integrated chatbot together;
2. Realtime online control of the robot arm;
3. An attention model selecting task-dependly relevant input information, influencing action and perception of the robot;
4. A behavioral system generating appropriate response behaviour given the sensory input and predefined behavioral dispositions ;
5. Robot motion generation inspired by biological motion avoiding repetitive patterns.

In many respects the entire research is still in its infancy, it is in progress as on the artistic side the Articulated Head is a work in progress, too. It will be continuously further developed: for instance, future work will include integrating a face recognition system and modelling memory processes allowing the Articulated Head to recall previous interactions. There are also already performances planned in which the Articulated Head will perform at different occasions with a singer, a dancer and its artistic creator. At all of these events the robot behaviour will be scripted as little as possible; the focus will be on interactivity and behaviour that instead of being fixated in few states *emerges* - emerges from the interplay of the robot's predispositions with the interactions themselves leading to a dynamical system that encompasses both machine and human. Thus, on the artistic side we will create - though only for the duration of the rehearsals and the performances - the situation we envisioned at the beginning of this chapter for a not too distant future: robots working together with humans.

12. References

Anderssen, R. S., Husain, S. A. & Loy, R. J. (2004). The Kohlrausch function: properties and applications, *in* J. Crawford & A. J. Roberts (eds), *Proceedings of 11th Computational Techniques and Applications Conference CTAC-2003*, Vol. 45, pp. C800–C816.

Bachiller, P., Bustos, P. & Manso, L. J. (2008). Attentional selection for action in mobile robots, *Advances in Robotics, Automation and Control*, InTech, pp. 111–136.

Bernstein, N. (1967). *The coordination and regulation of movements*, Pergamon, Oxford.

Bosse, T., van Maanen, P.-P. & Treur, J. (2006). A cognitive model for visual attention and its application, *in* T. Nishida (ed.), *2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2006)*, IEEE Computer Society Press, Hong Kong, pp. 255–262.

Breazeal, C. (2002). Regulation and entrainment in human-robot interaction, *The International Journal of Robotics Research* 21: 883–902.

Breazeal, C. & Scassellati, B. (1999). A context-dependent attention system for a social robot, *Proceedings of the 16th International Joint Conference on Artificial intelligence - Volume 2*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 1146–1151.

Brooks, A., Kaupp, T., Makarenko, A., Williams, S. & Oreback, A. (2005). Towards component-based robotics, *International Conference on Intelligent Robots and Systems (IROS 2005)*, Edmonton, Canada, pp. 163–168.

Burnham, D., Abrahamyan, A., Cavedon, L., Davis, C., Hodgins, A., Kim, J., Kroos, C., Kuratate, T., Lewis, T., Luerssen, M., Paine, G., Powers, D., Riley, M., Stelarc, S. & Stevens, K. (2008). From talking to thinking heads: report 2008, *International Conference on Auditory-Visual Speech Processing 2008*, Moreton Island, Queensland, Australia, pp. 127–130.

Call, J. & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later, *Trends in Cognitive Sciences* 12(5): 187 – 192.

Carpenter, M., Tomasello, M. & Savage-Rumbaugh, S. (1995). Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees, *Social Development* 4(3): 217–237.

Carruthers, P. & Smith, P. (1996). *Theories of Theories of Mind*, Cambridge University Press, Cambridge.

Castiello, U. (2003). Understanding other people's actions: Intention and attention., *Journal of Experimental Psychology: Human Perception and Performance* 29(2): 416 – 430.

Cavanagh, P. (2004). Attention routines and the architecture of selection, *in* M. I. Posner (ed.), *Cognitive Neuroscience of Attention*, Guilford Press, New York, pp. 13–18.

Cave, K. R. (1999). The FeatureGate model of visual selection, *Psychological Research* 62: 182–194.

Cave, K. R. & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search, *Cognitive Psychology* 22(2): 225 – 271.

Charman, T. (2003). Why is joint attention a pivotal skill in autism?, *Philosophical Transactions: Biological Sciences* 358: 315–324.

Déniz, O., Castrillión, M., Lorenzo, J., Hernández, M. & Méndez, J. (2003). Multimodal attention system for an interactive robot, *Pattern Recognition and Image Analysis*, Vol. 2652 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 212–220.

Driscoll, J., Peters, R. & Cave, K. (1998). A visual attention network for a humanoid robot, *Intelligent Robots and Systems, 1998. Proceedings., 1998 IEEE/RSJ International Conference on*, Vol. 3, pp. 1968–1974.

Emery, N. J., Lorincz, E. N., Perrett, D. I., Oram, M. W. & Baker, C. I. (1997). Gaze following and joint attention in rhesus monkeys (macaca mulatto), *Journal of Comparative Psychology* III(3): 286–293.

Faller, C. & Merimaa, J. (2004). Source localization in complex listening situations: Selection of binaural cues based on interaural coherence, *The Journal of the Acoustical Society of America* 116(5): 3075–3089.

Gerkey, B. P., Vaughan, R. T. & Howard, A. (2003). The player/stage project: Tools for multi-robot and distributed sensor systems, *International Conference on Advanced Robotics (ICAR 2003)*, Coimbra, Portugal, pp. 317–323.

Gielen, C. C. A. M., van Bolhuis, B. M. & Theeuwen, M. (1995). On the control of biologically and kinematically redundant manipulators, *Human Movement Science* 14(4-5): 487 – 509.

Heinke, D. & Humphreys, G. W. (2004). Computational models of visual selective attention: a review, *in* G. Houghton (ed.), *Connectionist Models in Psychology*, Psychology Press, Hobe, UK.

Herath, D. C., Kroos, C., Stevens, C. J., Cavedon, L. & Premaratne, P. (2010). Thinking Head: Towards human centred robotics, *Proceedings of 11th International Conference on Control, Automation, Robotics and Vision (ICARCV) 2010*, Singapore.

Herzog, G. & Reithinger, N. (2006). The SmartKom architecture: A framework for multimodal dialogue systems, *in* W. Wahlster (ed.), *SmartKom: Foundations of Multimodal Dialogue Systems*, Springer, Berlin, Germany, pp. 55–70.

Itti, L., Koch, C. & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(11): 1254–1259.

Kaplan, F. & Hafner, V. (2004). The challenges of joint attention, *in* L. Berthouze, H. Kozima, C. G. Prince, G. Sandini, G. Stojanov, G. Metta & C. Balkenius (eds), *Proceedings of the 4th International Workshop on Epigenetic Robotics*, Vol. 117, Lund University Cognitive Studies, pp. 67–74.

Kim, Y., Hill, R. W. & Traum, D. R. (2005). A computational model of dynamic perceptual attention for virtual humans, *14th Conference on Behavior Representation in Modeling and Simulation (brims)*, Universal City, CA., USA.

Kopp, L. & Gärdenfors, P. (2001). Attention as a minimal criterion of intentionality in robotics, *Lund University of Cognitive Studies* 89.

Kroos, C., Herath, D. C. & Stelarc (2009). The Articulated Head: An intelligent interactive agent as an artistic installation, *International Conference on Intelligent Robots and Systems (IROS 2009)*, St. Louis, MO, USA.

Kroos, C., Herath, D. C. & Stelarc (2010). The Articulated Head pays attention, *HRI '10: 5th ACM/IEEE International Conference on Human-Robot Interaction*, Osaka, Japan, pp. 357–358.

Kuhl, P. K., Tsao, F.-M. & Liu, H.-M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning, *Proceedings of the National Academy of Sciences* 100: 9096–9101.

Lakin, J. L., Jefferis, V. E., Cheng, C. M. & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry, *Journal of Nonverbal Behavior* 27: 145–162.

Liepelt, R., Prinz, W. & Brass, M. (2010). When do we simulate non-human agents? Dissociating communicative and non-communicative actions, *Cognition* 115(3): 426–434.

Metta, G. (2001). An attentional system for a humanoid robot exploiting space variant vision, *Proceedings of the International Conference on Humanoid Robots*, Tokyo, Japan, pp. 22–24.

Morén, J., Ude, A., Koene, A. & Cheng, G. (2008). Biologically based top-down attention modulation for humanoid interactions, *International Journal of Humanoid Robotics (IJHR)* 5(1): 3–24.

Ohayon, S., Harmening, W., Wagner, H. & Rivlin, E. (2008). Through a barn owl's eyes: interactions between scene content and visual attention, *Biological Cybernetics* 98: 115–132.

Peters, R. J. & Itti, L. (2006). Computational mechanisms for gaze direction in interactive visual environments, *ETRA '06: 2006 Symposium on Eye tracking research & applications*, San Diego, California, USA.

Pickering, M. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue, *Behavioral and Brain Sciences* 27(2): 169–226.

Riley, M. A. & Turvey, M. T. (2002). Variability and determinism in motor behaviour, *Journal of Motor Behaviour* 34(2): 99–125.

Saerbeck, M. & Bartneck, C. (2010). Perception of affect elicited by robot motion, *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 53–60.

Schneider, W. X. & Deubel, H. (2002). Selection-for-perception and selection-for-spatial-motor-action are coupled by visual attention: A review of recent findings and new evidence from stimulus-driven saccade control, *in* B. Hommel & W. Prinz (eds), *Attention and Performance XIX: Common mechanisms in perception and action*, Oxford University Press, Oxford.

Scholl, B. J. & Tremoulet, P. D. (2000). Perceptual causality and animacy, *Trends in Cognitive Sciences* 4(8): 299 – 309.

Sebanz, N., Bekkering, H. & Knoblich, G. (2006). Joint action: bodies and minds moving together, *Trends in Cognitive Sciences* 10(2): 70–76.

Shic, F. & Scassellati, B. (2007). A behavioral analysis of computational models of visual attention, *International Journal of Computer Vision* 73: 159–177.

Stelarc (2003). Prosthetic Head, New Territories, Glasgow, Interactive installation.

Stelarc, Herath, D., Kroos, C. & Zhang, Z. (2010a). The Articulated Head, NIME++ (New Interfaces for Musical Expression), University of Technology Sydney, Australia.

Stelarc, Herath, D., Kroos, C. & Zhang, Z. (2010b). The Articulated Head, SEAM: Agency & Action, Seymour Centre, University of Sydney, Australia.

Stelarc, Herath, D., Kroos, C. & Zhang, Z. (2011). The Articulated Head, Engineering Excellence Awards, Powerhouse Museum, Sydney, Australia.

Sun, Y., Fisher, R., Wang, F. & Gomes, H. M. (2008). A computer vision model for visual-object-based attention and eye movements, *Computer Vision and Image Understanding* 112(2): 126–142.

Tomasello, M. (1999). *The cultural origins of human cognition*, Harvard University Press, Cambridge, MA.

Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition, *Behavioral and Brain Sciences* 28: 675–691.

Ude, A., Wyart, V., Lin, L.-H. & Cheng, G. (2005). Distributed visual attention on a humanoid robot, *Humanoid Robots, 2005 5th IEEE-RAS International Conference on*, pp. 381 – 386.

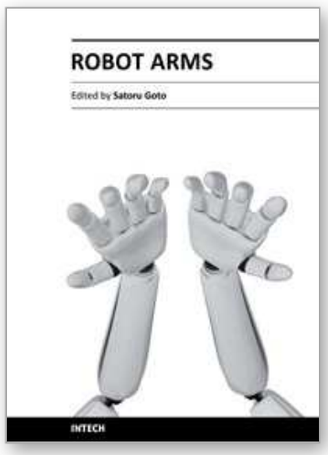
Wallace, R. S. (2009). The anatomy of A.L.I.C.E., *in* R. Epstein, G. Roberts & G. Beber (eds), *Parsing the Turing Test*, Springer Netherlands, pp. 181–210.

Wolfe, J. M. (1994). Guided search 2.0: a revised model of visual search, *Psychonomic Bulletin & Review* 1(2): 202–238.

Xu, T., Küandhnlenz, K. & Buss, M. (2010). Autonomous behavior-based switched top-down and bottom-up visual attention for mobile robots, *Robotics, IEEE Transactions on* 26(5): 947 –954.

Yu, Y., Mann, G. & Gosine, R. (2007). Task-driven moving object detection for robots using visual attention, *Proceedings of 7th IEEE-RAS International Conference on Humanoid Robots, 2007*, pp. 428–433.

IntechOpen



Robot Arms

Edited by Prof. Satoru Goto

ISBN 978-953-307-160-2
Hard cover, 262 pages
Publisher InTech
Published online 09, June, 2011
Published in print edition June, 2011

Robot arms have been developing since 1960's, and those are widely used in industrial factories such as welding, painting, assembly, transportation, etc. Nowadays, the robot arms are indispensable for automation of factories. Moreover, applications of the robot arms are not limited to the industrial factory but expanded to living space or outer space. The robot arm is an integrated technology, and its technological elements are actuators, sensors, mechanism, control and system, etc.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Christian Kroos, Damith C. Herath and Stelarc (2011). From Robot Arm to Intentional Agent: the Articulated Head, Robot Arms, Prof. Satoru Goto (Ed.), ISBN: 978-953-307-160-2, InTech, Available from: http://www.intechopen.com/books/robot-arms/from-robot-arm-to-intentional-agent-the-articulated-head

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen