2010

Thinking head: towards human centred robotics

Damith C. Herath
University of Western Sydney

Christian Kroos
University of Western Sydney,

Catherine J. Stevens
University of Western Sydney

Lawrence Cavedon
RMIT University

Prashan Premaratne
University of Wollongong, prashan@uow.edu.au

Publication Details

Herath, D. C., Kroos, C., Stevens, C. J., Cavedon, L. & Premaratne, P. 2010, 'Thinking head: towards human centred robotics', in C. Wen (eds), 2010 11th International Conference on Control, Automation, Robotics and Vision (ICARCV), IEEE, Singapore, pp. 2042-2047.

Thinking Head: Towards Human Centred Robotics

D. C. Herath*, Christian Kroos, Catherine J. Stevens
MARCS Auditory Laboratories
University of Western Sydney
Bankstown, Australia
*d.herath@uws.edu.au

Lawrence Cavedon
School of Computer Science and IT
RMIT University
Melbourne, Australia

Prashan Premaratne
School of Electrical, Computer and Telecommunications Engineering
The University of Wollongong
North Wollongong, NSW 2522, Australia

***Abstract*—Thinking Head project is a multidisciplinary approach to building intelligent agents for human machine interaction. The Thinking Head Framework evolved out of the Thinking Head Project and it facilitates loose coupling between various components and forms the central nerve system in a multimodal perception-action system. The paper presents the overall architecture, components and the attention system. The paper then concludes with a preliminary behavioral experiment that studies the intelligibility of the audiovisual speech output produced by the Embodied Conversational Agent (ECA) that is part of the system. These results provide the baseline for future evaluations of the system as the project progresses through multiple evaluate and refine cycles.**

***Keywords*—HRI, Perception, Cognition, Sensor Fusion, Dual Task**

I. Introduction

The "Thinking Head Project" is based on two stated goals, firstly, (i) to build a new generation Thinking Head embodying human attributes to improve human-machine interaction, and secondly, (ii) to build a plug-and-play research platform for users to test software in an interactive real-time environment. The goals complement each other in that, the plug-and-play research platform provides the necessary flexibility to evaluate various components that form the backbone of the 'thinking head' in a flexible manner allowing rapid evaluation and refining cycles to be performed leading to improved human-machine interactions.

The Thinking Head Framework (THF) evolved as a result of pursuing the aforementioned goals. A major challenge for the THF is to be flexible enough as an experimental platform to allow easy integration of components (or replacement of a component by one with similar capabilities), while supporting processing speeds required for real-time behaviours, including the control of robotic components. Given the nature of the TH project - for which audio-visual input processing and rendering is a critical focus -- efficient processing of multiple high-volume data streams is a particular challenge.

In the first half of the paper we outline the design and associated decisions made in implementing the THF, and describe its utility in integrating a number of human-robot interaction (HRI) capabilities. The most novel of these is a central behavioral unit that models "attention" to a human interaction participant. This unit drives the actions of the robotic framework in its HRI setting and therefore requires highly efficient real-time performance.

In the second half, we present the first evaluation study of the system using the dual task paradigm [1], a method in cognitive psychology adapted to infer the relative intelligibility of the audiovisual speech synthesis generated by the Embodied Conversational Agent (ECA) that is the computer graphics humanoid front-end of the system.

The paper is organized as follows. In section II our software integration framework is discussed. Section III describes the currently implemented Human-Machine interfaces. Together sections II and III address the second set of stated goals in the Thinking Head Project. Section IV presents the attention and behavioral system which encompass the stated goal of embodying human attributes to improve human-machine interaction. In section V, the first evaluation cycle of the system is presented with results and an interpretation of the results. Section VI concludes the paper with a summary of planned future work.

II. Integration Framework

The integration framework for our system combines approaches from open agent-oriented systems previously used for multimodal dialogue systems (e.g., [2, 3]) and frameworks for high-performance robotic platforms (e.g., [4]). 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 real-time 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 (e.g., as used in [5]) which incur overheads on message-passing to components. The CoSy Architecture [6] shares similar motivations and characteristics. Fig. 1 shows the integration framework.

A. Event Driven Middleware

In common with other dialogue platforms, we use an event-driven framework, which has a number of desirable properties, such as: naturally modeling 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 optimizing the system, via inter-component configuration commands for particular interaction states.

We use the Boost library for underlying TCP/IP support. A dynamically configurable shared memory architecture is also provided for high frequency/high bandwidth applications such as those requiring streaming data.

B. Software Interface

The framework supports components written in multiple languages running on diverse software platforms; it also allows multiple versions of similar-type components, with a policy for selecting contributions from components to be specified. For example, the system may contain two dialogue managers, with a "dialogue event" being sent to both, with each dialogue manager processing that event and suggesting a response. The selection policy chooses amongst the responses.

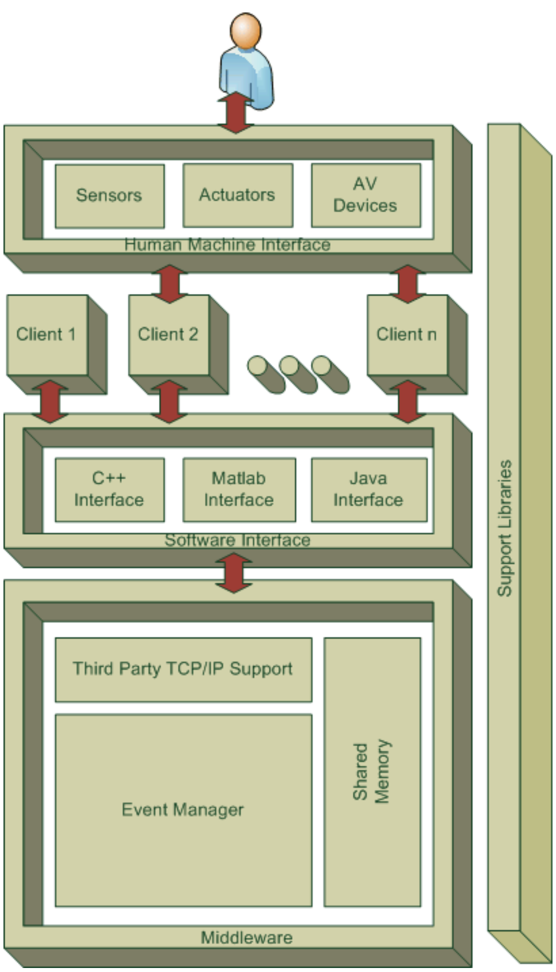

Fig. 1. Thinking Head Integration Framework

C. Human-Machine Interface

The human-machine interaction is realized through the various hardware devices integrated within the framework. These include various sensor, actuator and audio visual devices. These will be discussed in detail in the next section.

III. Human Robot Interfaces

The Human-Robot interactions are realized through the various sensors, actuators and audio visual devices available within an implementation of the framework. Following components are currently available and are used as appropriate in various configurations (Section V describes two different configurations).

A. Auditory Localisation

The auditory localization system provides accurate information on the instantaneous locations (azimuth) of multiple moving interlocutors in a noisy and reverberant environment. Localization is limited to the half sphere in front of the TH agent and provides azimuth angle from about -90° to +90°. The azimuth system uses a microphone-pair mounted in front of the users. The localization is based on Faller and Merinaa [7] which has been modified and adjusted to the thinking head setup.

A measurement of the coherence between the different microphone signals is also added to each instantaneous localization value. Coherence provides a measure of how similar the different microphone signals are, and provides an indirect measure of the disturbance by background noise and room reverberation. Therefore, the coherence provides useful information to the higher levels of the Agent on the reliability of instantaneous localization measures. During speech the coherence is rather high ($c12 \approx 0.7$-1) and during speech pauses the ambient background noise typically produces rather low values ($c12 \approx 0.2$-0.4). Due to this property of the coherence, a simple threshold device can be implemented that only considers instantaneous localization values if the corresponding coherence value is above a predefined threshold (e.g., $c12{,}limit = 0.6$). In fact more sophisticated methods based on the coherence are used in the higher level decision making process of the Attention Model.

B. Visual Tracking

We have adopted two commercially available systems for tracking people in 3D and faces in close proximity.

The 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. The tracking system is capable of tracking multiple persons with considerable tolerance to occlusion and occasional disappearance from the field of view.

The face tracking algorithm is capable of detecting a single face in the camera's field of view and then continuously tracking the detected face with a high degree of accuracy withstanding considerable occlusion, scale variance and deformations.

C. Gesture Recognition

Hand gesture recognition system can be an important step in effective communication between a human and a robot. A system was developed with a high precision real time capability consisting of 10 unique hand gestures to effectively communicate with a computer interface [7]. The system known as the 'Consumer electronics control system using hand gestures' is a new innovative user interface that resolves the complications of using numerous remote controls for domestic appliances. Based on one unified set of hand gestures, this system interprets the user hand gestures into pre-defined commands to control aspects of the robotic system. The system has been tested and verified under natural, incandescent and fluorescent lighting conditions.

D. Prosthetic Head

Prosthetic Head refers to a software 3D animated head displayed on a LCD screen, (Fig. 2) an Embodied Conversational Agent (ECA). The visual front-end of the Thinking Head is a three-dimensional computer-graphic representation of a human face which is capable of visual speech movements and of displaying basic emotional expressions.

E. Articulated Head

The Articulated Head (AH) [8, 9] was primarily conceived as a work of art by the artist Stelarc (Fig. 3) extending the original Prosthetic with a robotic embodiment. The robot arm has six degrees of freedom of movement, but is mounted at a fixed location. In this configuration the framework utilize all the interfaces discussed in section III to provide the human participant with an engaging interactive experience. It provides an innovative solution for an embodiment for an AI (Artificial Intelligence) agent that is aesthetically pleasing. It was designed not only to better embody a software agent and produce a stronger sense of presence but also to produce a more emotive and artistic performative installation.

From a HRI perspective, Articulated Head, therefore is a significant step towards the evaluation of more complex human-robot interactions. It allows a multiplicity of different ways to research these different aspects. The LCD screen mounted on the end of the 6 degree-of-freedom industrial robot arm effectively becomes the neck of the AI agent. The advantage of this configuration is that the virtual behaviour of the Prosthetic Head can be augmented, counterpointed or synchronized with the motion of the robot. The robot therefore allows us to have a library of articulated movements of the LCD screen to turn CW or CCW, bend forwards and backwards, nod up and down, and swivel from side to side. The industrial robot arm provides precision and robustness with a variation of speed from imperceptibly slow to very fast motion allowing the programming from subtle and gentle to quite aggressive gestures.

IV. The Thinking Head Attention and Behavioural System (THAMBS)

The human-robot interfaces described in the previous section provide the AH with information about its environment. The obtained data exhibit an interesting conceptual dichotomy with respect to the higher processing levels of the AH. On the one hand, as a representation of the rich natural environment of the AH, they are very sparse picking up only a few important aspects of the surroundings. On the other hand, when faced with the task of generating a behavioural response to them, they are already too rich and complex to be able to react to them with a reflex loop or a single stream stimulus-response mechanism, in particular, when considering their enfolding over time (e.g., acknowledging the difference between a person approaching the AH quickly or slowly).

Thus, the behavioural system steering the AH must be able to cope with incomplete information and still be selective about the information that are forwarded to processes which generate the behavioural response. An attention model directly addresses the latter and indirectly the former as it makes the lack of completeness explicit for the higher processing stages, i.e., these stages never have access to all information and need to actively acquire relevant additional bits if deemed necessary.

Fig. 2. Prosthetic Head

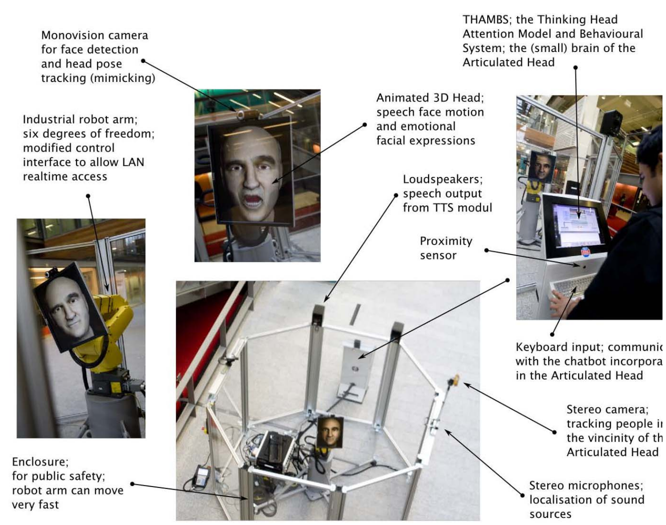

Fig. 3. Articulated Head

For instance, if an interacting person's face fails to be tracked by the FaceAPI tracker, the AH can attempt to bring its monovision camera closer to the person. This might seem quite obvious, but the fact that the computer science fields active vision and active hearing/listening have emerged relatively recently, appears to tell a different story, the story of perception (human, animal or machine) being considered passive information intake and decoupled from action, that is, from the motor system. Psychological theories such as the Theory of Event Coding (also known as Common Coding Theory) [10] and Ecological Psychology [11] as well as, specifically for speech, the Motor Theory of Speech Perception [12] have challenged this perspective - though for different reasons - and proposed a strong link between action and perception. Based on this school of thought even a computational framework has been suggested linking motor control and social interaction [13].

We argue therefore that consistent interactions with humans emerge only if the robot's sensing capabilities of the environment are related to its motor capabilities in a meaningful way and according to the expectation of the human user. Only a tight coupling between perception and action can generate behaviour that convincingly creates the illusion that the AH is an intentional agent with its own agenda and with this enable a different quality of human-robot interaction. The Thinking Head Attention and Behavioural System (THAMBS) was developed against this background.

THAMBS is a perception-action control architecture that consists of the following high-level modules: (1) a perceptual system, (2) an attention system, (3) a central control system and (4) a motor system.

The perceptual system wraps the lower level sensing streams and creates within-system standardised perceptual events. Currently three sensing abilities of the AH are integrated: acoustic source localization, visual people detection in 3D space, and face tracking. The perceptual system has its own set of thresholds acting for instance on the confidence values returned by the sensing systems and also computes the deltas for each input (‘velocities’).

The generated perceptual events are passed on to the attention system. Algorithmic attention models have been studied for some time. The majority of them are biologically inspired [14-16] and have been only applied in computer agents acting in a virtual environment [17] bypassing the extremely difficult task of real world object recognition (but see e.g., [18]). The identity of objects placed in a virtual environment can directly made known to the attention model of the agent; an option that is clearly not available when dealing with a robot and real world sensing.

The attention system of the AH checks the generated perceptual events individually against attention thresholds specific to each type of perceptual event. Those events that have values below the threshold are considered ‘subliminal’ and can be still further processed but will never be fed to the central control system. Note that there are thresholds for many aspects of the perceptual event, e.g., a high velocity of an otherwise sub-threshold event can allow it becoming attended. Perceptual events that pass the test will create an attention focus that currently is entirely spatially organized (“pay attention to region X”), thus, identity of two foci is assumed if they refer to the same spatial region. In the future semantic criteria will be introduced on top of the spatial mapping.

The attention system then assigns an initial weight and an exponential decay function to the focus based on the current task priorities specified by the central control system. These depend, of course, on the overall state of the AH with respect to the ongoing interaction and its ultimate goal. The attention system determines a single attended event from all available foci using a winner-takes-all strategy and relays it to the central control system as the presently attended event. It also directly generates a motor goal to bring the attended event at the centre of the AH’s mobile visual system and forwards this motor goal directly to the motor system.

The central control system evaluates the attended event based on the values of a larger set of THAMBS state variables and generates a behavioural response, i.e., a pre-defined temporal sequence of motor goals. The behaviour trigger is currently realised in form of conditional rules acting on various thresholds. In the future, however, we will include very simple simulations of cognitive and affective processing that will evaluate the attended event according to its distance to potential behavioural responses in a multidimensional parameter space spanning among other parameters state variables that characterise the current affective states of the AH. If the attended event is sufficiently close (varying thresholds) to one of the pre-defined behavioural responses, this behaviour i.e., a temporal sequence of motor goals, will be considered an appropriate response and activated.

The abstract motor goals (e.g., “follow person with id 2”) will be transformed into sequences of implementation-specific motor primitives by the motor system. The set of motor primitives covers both movements of the robot arm and facial movements of the ECA displayed on the monitor. Motor goals coming from the central control systems will suppress goals from the attention system, unless the latter have an associated weight higher than task-specific threshold.

V. Evaluating an ECA Using a Dual Task Experiment

A. *Background, Hypotheses, and Method*

To maintain rigorous empirical testing and produce interpretable results, psychological research methods dictate control or randomisation of extraneous variables not part of the systematic experimental manipulations. As a consequence, evaluation methods have to evaluate partial systems first to establish baselines and then add stepwise complexity, with regard to both, the system under investigation and its environment.

We are currently developing new methods for the evaluation of specifically the ECA used in the system. Research in cognitive psychology provides evaluation approaches that can be adapted to HRI, are rigorous and have been proven to reliably measure aspects of cognitive processes of the participant in a controlled experimental environment.

One such new evaluation method is based on a dual task paradigm. The paradigm involves performing two tasks concurrently resulting in impaired behavioural performance on one or both tasks [1, 19]. The dual task paradigm allows assessing the cognitive load of the primary tasks by forcing the participant to divide attention across two tasks whereby the secondary task is chosen to enable a straight-forward quantification of the degree of interference from the primary task,e.g., by using response time measurements.

In a recent evaluation experiment, participants performed a cognitive word-based primary task and secondary reaction time (RT) task at the same time. The primary task had two levels of difficulty. The easy version involved shadowing or saying aloud the word that was uttered by the ECA – the spoken word being a sensory cue. The more difficult version of the primary task required the participant to name the superordinate category to which the word belonged – in this case the spoken word is a semantic cue. In terms of a flexible view of attention, relatively early selection (shadow the word) is possible with a sensory cue but a later mode of selection (categorise the word) is necessary when the word serves as a semantic cue.

The secondary task required a button press response to a visual target on the ECA's face; the target was a small fly. The secondary task was used to measure potential capacity expended on the cognitive task. The rationale is that the greater the capacity allocated to the cognitive task the less capacity available for monitoring the fly and the longer the RTs on the secondary task should be [20-22].

Using this basic dual task paradigm, we compared the facilitation or impediment on processing achieved by the presence of an ECA producing the primary task sensory or semantic cues. In the auditory-visual (AV) condition, the ECA uttered individual word items and a participant saw the ECA utter the words. In the auditory only (A) condition, the ECA was present but there were no lip movements, only the voice uttering the individual word items.

It was hypothesized that if the ECA AV model is effective and intelligible then this should facilitate shadowing and we should see equal or reduced RTs on the secondary task in the AV versus A condition. Conversely, if the AV model is not effective then there will be no difference or possibly poorer secondary task RTs on the AV versus A conditions. The relatively demanding category naming task was included to investigate any interaction between primary task demand and multi versus unimodal stimuli on secondary task RTs. A baseline of RTs on the fly swatting task was obtained by presenting the secondary task on its own. This serves as a reference from which to measure the capacity (RT) required for the cognitive task. The secondary task RT ordering should be: baseline < shadowing < category naming. The relative intelligibility of the speech model can be gauged from shadowing accuracy. Differences in accuracy across modality and tasks are not anticipated. Accuracy on the secondary RT task will reflect vigilance on that task. Self-report ratings of ECA likeability, engagement, etc., were also obtained.

The sample consisted of 40 female undergraduate students (Mean age=20.6 years, SD=6.42) who completed the experiment for partial course credit. Twenty participants were assigned to the AV condition and 20 to the A condition; participants performed baseline (single task), shadowing and category naming tasks in counterbalanced orders so as to distribute serial order effects.

TABLE I

Mode Ratings of ECA and Interaction Quality, Enjoyment and Engagement; minimum possible rating is 1 ("totally disagree") and maximum possible rating is 5 ("totally agree").

Item	A-Only	AV
I find the Head likeable	4	4
I find the Head engaging	4	4
I find the Head easy to understand	2	2
I find the Head life-like	5	4
I find the Head humorous	3	4
The Head kept my attention	4	4
I would like to interact with the Head again	3	4
I enjoyed interacting with the Head	3	4
I felt as if the Head was speaking just to me	5	5

B. *Results*

The secondary task (fly swatting) mean RT in the baseline (single task) condition was 429.13 ms (SE=3.45). As hypothesized, when the primary task was also performed (dual task), RTs were significantly faster while shadowing (M=581.80, SE=4.58) than while category naming (M=672.36, SE=5.58), F(2, 2254)=845.28, p<.001, η^2_p=.43. There was a significant interaction of word task with modality such that RTs were longer in the AV condition compared with the A condition especially while shadowing, F(2, 2254)=7.11, p=.001, η^2_p=.006.

Performance on the fly swatting task showed significantly greater accuracy recorded in the baseline condition (M=99, SE=.003) followed by shadowing (M=.99, SE=.005) then category naming conditions (M=.96, SE=.009), F(2,37)=5.80, p=.006, η^2_p=.24. Accuracy was >95% indicating vigilance on the secondary task was very good.

Latencies recorded from shadowing (primary task) were not affected by modality of stimulus presentation. The mean shadowing latency in the A condition was 372.52 ms (SD=158.82) and the AV condition it was 366.14 ms (SD=151.48).

As hypothesized, performance on the primary task showed that category naming (M=.86, SD=.08) was significantly more difficult than shadowing (M=.91, SD=.03), F(1,38)=13.68, p=.001, η^2_p=.27. There was no significant word task x modality interaction.

Table 1 shows the mode (highest frequency) self-report ratings assigned to the nine rating scale items for Auditory-only and Auditory-Visual conditions. The results of t-tests conducted on the ratings indicate that the mean ratings all differ significantly from the midpoint of the scale (3: neither agree nor disagree) for both auditory only t(8)=30.61, p<.001 and for AV conditions t(8)=33.16, p<.001; ratings did not differ significantly from each other. A one-way between-subjects ANOVA showed no effect of modality on mean ratings, A-only (Mean=3.66, SD=0.36) and AV (Mean=3.52, SD=0.32).

C. Discussion

Results of evaluation involving a relatively primitive ECA model shows that the AV speech model does not enhance user perception. In fact under some circumstances, when task demand is high and the concurrent task relies on speech perception, e.g., shadowing, performance in response to the current AV model impedes RT relative to the auditory only condition. Performance on the primary task, reflected in shadowing accuracy and latency, is not affected by modality with comparable results in AV and auditory only conditions.

VI. Conclusion

We have presented an open HRI framework that enables easy re-configuration and addition/removal of components for various instantiations of robotic hardware, sensing interfaces, computer graphic animations and control software centered around an embodied conversational agent, the Thinking Head. Our current configuration, the Articulated Head has multiple sensors that provide enhanced situational awareness enabling it to react to changes in the ambience through a range of engaging motor actions. In line with the requirement of flexibility throughout the framework, the perception-action control system of the AH, THAMBS, generates behavioural responses to user actions on an implementation-independent abstract level employing modular perception and motor system representations as well as an attention model for organising and prioritising aspects of the incoming information according to the current task.

We also showed that a dual-task evaluation paradigm is able to uncover performance short-comings that would remain unnoticed using traditional questionnaire-based evaluation methods.

Future work will include testing the flexibility of the approach with substantially different configurations (e.g., a mobile robot) and developing a comprehensive evaluation metric consisting of a series of individual tests that can be integrated to evaluate an overall complex system such as the Articulated Head presented here.

Acknowledgment

The Thinking Head project is funded by an ARC/NHMRC special initiative on Thinking Systems. Authors would also like to acknowledge the contributions made by Zhengzhi Zhang, Yvonne Leung, and Guillaume Gibert. Also the authors would like to thank Stelarc for insightful discussions with him, his artistic interpretations and most of all providing the impetus for the project through his prosthetic head (thanks also go to the original developers) and the Articulated Head.

References

[1] C. Karatekin, J. W. Couperus, and D. J. Marcus, "Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses," *Psychophysiology,* vol. 41, pp. 175-185, 2004.

[2] G. Herzog and N. Reithinger, "The SmartKom Architecture: A Framework for Multimodal Dialogue Systems," in *SmartKom: Foundations of Multimodal Dialogue Systems*, 2006, pp. 55-70.

[3] A. Cheyer and D. Martin, "The Open Agent Architecture," *Journal of Autonomous Agents and Multi-Agent Systems,* vol. 4, pp. 143-148, 2001.

[4] "Orca Robotics Open Source Framework (http://orca-robotics.sourceforge.net)." 2010.

[5] M. Schroder, "The SEMAINE API: Towards a standards-based framework for building emotion-oriented systems," *Advances in Human-Computer Interaction,* vol. to appear, 2010.

[6] N. Hawes, J. L. Wyatt, M. Sridharan, H. Jacobsson, J. R. Dearden, A. Sloman, and G.-J. Kruij, "Architecture and representations," in *Cognitive Systems Research*, H. I. Christensen, Ed. Berlin: Springer, 2010, pp. 51-93.

[7] P. Premaratne and Q. Nguyen, "Consumer electronics control system based on hand gesture moment invariants," *IET Computer Vision,* vol. 1, pp. 35-41, 2007.

[8] C. Kroos, D. C. Herath, and Stelarc, "The Articulated Head: An Intelligent Interactive Agent as an Artistic Installation " in *Workshop on Performance Evaluation and Benchmarking for Next Intelligent Robots and Systems at IEEE/RSJ IROS 2009*, St.Louis, MO, USA, 2009.

[9] C. Kroos, D. C. Herath, and Stelarc, "The Articulated Head Pays Attention," in *5th ACM/IEEE International Conference on Human-Robot Interaction* Osaka: ACM/IEEE, 2010.

[10] B. Hommel, J. Musseler, G. Aschersleben, and W. Prinz, "The theory of event coding (TEC): A framework for perception and action planning," *Behavioral and Brain Sciences,* vol. 24, pp. 849-878, 2001.

[11] J. J. Gibson, *The ecological approach to visual perception*, 1986 (Original work published 1979).

[12] A. M. Liberman and I. G. Mattingly, "The motor theory of speech perception revised," *Cognition,* vol. 21, pp. 1-36, 1985.

[13] D. M. Wolpert, K. Doya, and M. Kawato, "A Unifying Computational Framework for Motor Control and Social Interaction," *Philosophical Transactions: Biological Sciences,* vol. 358, pp. 593-602, 2003.

[14] R. J. Peters and L. Itti, "Computational mechanisms for gaze direction in interactive visual environments," in *Proceedings of the symposium on eye tracking research & applications*, San Diego, California, 2006

[15] T. Bosse, P.-P. v. Maanen, and J. Treur, "A Cognitive Model for Visual Attention and Its Application," in *Proceedings of the IEEE/WIC/ACM international conference on Intelligent Agent Technology*: IEEE Computer Society, 2006.

[16] Y. Sun, B. Fishe, H. Wang, and M. Gomes, "A Computer Vision Model for Visual-Object-Based Attention and Eye Movements," *Computer Vision and Image Understanding,* vol. 112, pp. 126-142, November 2008 2008.

[17] Y. Kim, R. W. Hill, and D. R. Traum, "A Computational Model of Dynamic Perceptual Attention for Virtual Humans," in *14th Conference on Behavior Representation in Modeling and Simulation (BRIMS)* Universal City, CA, 2005.

[18] C.Breazeal and B. Scassellati, "Context-dependent attention system for a social robot," in *International Joint Conference on Artificial Intelligence*, 1999, pp. 1146 - 1153.

[19] D. D. Salvucci and N. A. Taatgen, "Threaded cognition: an integrated theory of concurrent multitasking," *Psychological Review,* vol. 115, pp. 101-130, 2008.

[20] W. A. Johnston and S. P. Heinz, "Flexibility and capacity demands of attention," *Journal of Experimental Psychology: General,* vol. 107, pp. 420-435, 1978.

[21] D. Kahneman, *Attention and effort*: Englewood Cliffs, NJ: Prentice-Hall, 1973.

[22] C. D. Wickens, "Multiple resources and performance prediction," *Theoretical Issues in Ergonomic Science,* vol. 3, pp. 159-177, 2002.