

Robotic Etiquette: Structured Interaction in Humans and Robots

Bernard Ogden and Kerstin Dautenhahn*

Adaptive Systems Research Group
Department of Computer Science
University of Hertfordshire
College Lane, Hatfield, Herts AL10 9AB
United Kingdom

Abstract. Visual nonverbal behaviour plays a significant role in allowing humans to structure their interactions meaningfully, thus facilitating communication between individuals. Such behaviour may be exploited in the design of socially intelligent robots. While much visual behaviour (e.g. facial expression) is hard to detect using a machine vision system, especially in unconstrained situations such as natural interaction, simple gestures and the positions of interactants are relatively easy to detect. This paper begins with a description of some related work, followed by an overview of some systems that track and/or interpret human movement. This is followed by a discussion of what is known about the structure of human interactions and a detailed example of a particular kind of structured interaction (the greeting). Ways in which interactional structure can be exploited in the development of socially intelligent robots are considered: finally, a specific application involving the use of robots in the rehabilitation of autistic children is described. We refer to this approach, using knowledge of interactional structure both to generate social behaviour for a robot and to interpret interactions for a machine vision system, as *interactive vision*. The work described in this paper is still in its early stages and there are no experimental results to report as yet: the purpose of this paper is rather to outline ways in which sociological and psychological work on human interactional behaviour can be exploited in socially intelligent agents and machine vision systems.

In designing robotic agents to interact with humans it is important that the robot be able to determine what the human is doing and respond appropriately. A number of features of human behaviour are likely to contain useful information: the most obvious of these is speech, which is our primary channel for information exchange and which also contains a significant amount of paralinguistic content which serves, for instance, to convey information about mood, alter the meaning of words and sentences (for instance using a ‘questioning’ tone) and to structure interactions by helping to indicate when an individual is coming to the end of their turn at speaking.

Other forms of nonverbal behaviour also convey information that is useful in carrying out interactions. Much work has been done on visual nonverbal behaviour, including such features as facial and gestural expression and use of space (e.g. Collett, 1983; Hall, 1966; Kendon, 1980; Kendon, 1990; Schefflen, 1973; Ekman and Friesen, 1969 and Gill et al, 1999). It has been argued that visual nonverbal behaviour plays a significant role in allowing humans to structure their interactions meaningfully, making communication possible, as well as conveying information about emotion and affecting the meaning of linguistic communication in a similar way to the paralinguistic content of speech. Furthermore, while much of this information (e.g. facial expression) is hard to detect in unconstrained situations such as natural interaction, simple gestures and especially information about the use of space are relatively easy to detect.

Human interactional behaviour possesses a structure (Kendon, 1980; Kendon, 1990; Schefflen, 1973) that can be exploited in the design of socially intelligent agents, in order both to understand the behaviour of other agents and to generate behaviours for the agent. A number of systems that recognise and / or engage in human-like interactive behaviour have already been developed but these systems tend not to make use of the principles of structured interaction described in the psychological and sociological literature. For example, Oliver, Rosario and Pentland (1999) describe a vision system to recognise interactions in a surveillance task using coupled hidden Markov models (Brand, 1997). This work depends on statistical learning techniques to recognise interactions and does not employ explicit a priori knowledge about interactional structure.

The humanoid robotic head, Kismet, behaves in a human-like manner in order to engage humans and to make its behaviour easier to understand, for example by showing human facial expressions and gaze patterns (Breazeal and Fitzpatrick, 2000). Unlike the previously described work, Kismet generates interactive behaviour rather than attempting to observe it. Kismet provides an example of how human-like social behaviour can be exploited by an artefact although, again, structured interactional principles are not involved in this work.

Johnson, Galata and Hogg (1998) describe a system that learns interactional behaviours from observations of human interactions. This system observes interactions and then attempts to engage in them. The interactions

* {B.Ogden | K.Dautenhahn}@herts.ac.uk

involved are quite simple (in this case a handshake, which would be just one component in a larger greeting interaction) and principles of structured interaction are again not employed: the system functions by replicating observed behaviours. However, this is related to the present work in that it involves an agent both observing and attempting to engage in interaction.

Benford and Fahlén (1993) provide an example of work that does use principles that have been described in the human sciences. They describe the use of a spatial model of interaction in collaborative virtual environments in order to facilitate interaction. The principles described are very similar to those from the human sciences literature on proxemics (the social use of space): this work is thus an example of the usefulness of applying such principles in computational systems, although the agents in this case are avatars of humans rather than autonomous artificial agents.

1 Machine Vision for Human Movement Tracking

A number of approaches to the tracking of human movement for machine vision systems have been developed, employing techniques varying from the recognition of action directly from the image sequence (Davis and Bobick, 1997) to the construction of detailed deformable models (Kakadiaris and Metaxas, 1998). (Essa (1999) and Aggarwal and Cai (1999) provide reviews of machine vision systems for analysing human movement, while Crowley (1997) discusses the use of machine vision for the development of more natural interfaces for man-machine interaction.) To be useful in a robot that is to interact with real people in real settings the vision system itself must obviously operate in real time and be robust enough to deal with problems such as partial occlusion and recognition of actions from multiple angles. Obviously simpler systems have a significant advantage when it comes to operating in real time but they may also be less robust than more complex systems and will certainly be unable to extract similarly detailed information from the image sequence.

At the simplest end of the spectrum are systems that recognise actions directly from image sequences without constructing models (Davis and Bobick, 1997; Seitz and Dyer, 1997). Such systems may be quite restricted in the kinds of actions that they can recognise (e.g. restricted to cyclic motions) and may have significant problems with self-occlusion and incidental motion (motion that occurs along with an action but is irrelevant to that action, such as the walking motion when an action is performed by someone who happens to be walking at the same time (Davis and Bobick, 1997)).

More sophisticated action recognition becomes possible when a model is constructed: this allows actions to be recognised from the model rather than directly from the image sequence. Incidental motion becomes less of a problem as actions can be defined in terms of the relevant parts of the model. Many model-based systems can also infer the location of body parts that are currently out of sight, thus dealing with problems of occlusion. As action recognition is based on the model rather than directly on the image sequence differing view angles also become less of a problem: as long as the model can be accurately constructed from different angles actions can be recognised. Also, three dimensional models can be constructed to allow recognition of a wider range of actions.

Some model-based systems are very simple e.g. Azarbayejani, Wren and Pentland (1996), producing only a very approximate model of the human body (e.g. different blobs representing limbs, torso and head). Such a model may be adequate for locating a person in a scene and recognising a limited range of simple gestures (e.g. arms out at side vs. arms above head) but is certainly not capable of recognising complex gestures and would not be suitable for a vision system intended to interpret body language. A little more complex are systems employing articulated but two dimensional models e.g. Ju, Black and Yacoob (1996). Such systems can be capable of recognising reasonably complex motions: however, human body movement is three dimensional so it seems that such systems may be inherently limited. Furthermore, two dimensional approaches are likely to have problems with self-occlusion (Gavrila and Davis, 1995).

At the high end of the spectrum of complexity are systems employing three dimensional articulated models (e.g. Kakadiaris and Metaxas, 1998; Wren and Pentland, 1998). Such systems should, at least in principle, be capable of recognising the full range of human body movement. These systems may employ a simple skeletal model, or they may use deformable models to represent the whole of the human body. However, the latter approach may be of limited value in the recognition of body movements as, with a few exceptions (e.g. in the face), the detailed deformations of the human body seem to contribute relatively little to body language. Also, such deformations are often not visible in natural interactions as interactants are usually clothed. Therefore such approaches, with their increased computational demands, seem unnecessary in applications dealing with visual nonverbal behaviour. Three

dimensional skeletal models would be ideal, but again the computational demands are high and the existing state of the art is still quite limited for dealing with natural situations (e.g. Wren and Pentland (1998) only deal with the upper body of a seated human). Thus, with limited computational resources and a requirement of operating in real time, approaches using simpler two dimensional models or blob models may be most appropriate despite their restricted ability to extract detailed information. In any case, as we will discuss, just knowing the positions and orientations of interactants may be very useful.

2 Structured Interaction

Human interaction has a definite structure that has been hypothesised to operate as a kind of program followed by the interactants: furthermore, a number of nonverbal metacommunicative behaviours are involved in describing and maintaining the structure of an interaction (Kendon, 1990; Schefflen, 1973). While human communication involves all of the sensory modalities (Birdwhistell, 1970; Hall, 1966), many of the behaviours that have been associated with interactional structure occur in the visual modality. Such behaviours may be divided into two broad types: kinesic (communicative body movements such as gestures) and proxemic (involving the relative positioning and orientation of interactants).

Each action in an interaction is hypothesised to occur in a slot, with the meaning of an action depending on the slot in which it occurs and the completion of an action setting up a new slot. For example, if a waitress approaches a man in a coffee shop with an inquiring expression then a slot in which he may order coffee is created. Alternatively, if she approaches him and starts wiping the table then this slot is not created. Equally, should the waitress use the same facial expression in a different context its meaning might be interpreted slightly differently: in a conversation with friends it would likely be interpreted as a request for further information about something and certainly wouldn't be interpreted as a request for an order (Kendon, 1980).

Essentially this notion of programmatic structure governing interaction corresponds to the notion of behaviour being organised by scripts or schemata (Schank and Abelson, 1977). It should be emphasised, though, that this 'programmatic' structure does not impose a strict set of rules on encounters – humans are noted for their tendency to act unpredictably. Variation, though, is often a part of the program: a program is likely to possess a set of alternative behaviours that may fill any given slot and truly idiosyncratic and novel behaviours may be quite rare (Schefflen, 1973). The structure of the program itself may also vary in certain circumstances: sometimes stages in an interaction may be skipped (as in the case where people who have already greeted each other once at a given event meet again) or a single behaviour may sometimes accomplish more than one stage at a time (Kendon, 1980).

Purely visual information may be used to at least some extent to identify types of interaction and changes in stages of interactions. It is useful to distinguish between proxemic and kinesic factors if only because proxemic information is more easily obtained by a machine vision system.

From a proxemic point of view, interactions (indeed, any activity) occur with a particular space which may be referred to as a transactional segment. For example, a person watching television will have a transactional segment covering the space between where he is sitting and the television: other people will likely hesitate to cross this space, as it belongs to the actor. Interactants form a mutually shared 'o-space' consisting of their overlapping transactional segments. Creation and maintenance of interactional spaces is considered to play an important role in establishing the joint attention that is necessary for an interaction to occur: any interactional space should differ sharply from the arrangements that precede and follow it. Also, it should be possible for an interactant to propose a new kind of interaction by moving to a new position and thus changing the space: the other interactants can then either allow the change or reject it by shifting in such a way as to cancel out the change (Kendon, 1990). An interactional space may also be associated with particular objects: for example, people playing a card game will usually be seated at a table (Kendon, 1980). (The association of certain kinds of action with particular kinds of object has already been exploited in machine vision (Moore, Essa and Hayes, 1999).)

The above considerations lead to the concept of spatial-orientational shifts (Kendon, 1990), hereafter referred to as S/O shifts. This phenomenon is particularly useful for our purposes. Because interactions are located in a particular space S/O shifts may be seen as markers of phases in interactions – a new phase in an interaction must be marked by a shift in the relative positions and orientations of the interactants, although the shift may sometimes be very subtle (e.g. involving just a redistribution of weight of the movement of a foot (Kendon, 1990, p192). The distance between interactants may also be indicative in itself of the type of interaction that they are engaged in: in

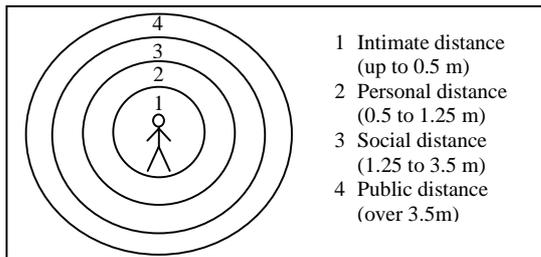


Figure 1: Significant distances in American interaction. Each distance may be split into a 'near' and 'far' phase and exact ranges may vary with context.

American culture four basic distances associated with different kinds of interaction have been identified (Hall, 1966): see figure 1.

Kinesic information can also be useful in identifying the structural features of an interaction. For instance, a 'head dip' gesture has been identified as possibly indicating a change in focus of attention (Kendon, 1990). In addition to this, particular gestures are associated with particular phases of particular interactions: turn-taking in conversation, for example, tends to be associated with postural shifts (Kendon, 1990) and certain kinds of body moves (Gill et al, 1999). Thus, identification of a gesture will give some indication of the kind

of interaction that is occurring, as well as the current phase in that interaction. It should be noted, however, that some gestures are quite idiosyncratic and may only have meaning given knowledge about the person performing the act (Ekman and Friesen, 1969).

3 Greetings: An Example of Structured Interaction

The structure of greeting interactions is described in some detail by Kendon (1990). This section summarises the stages involved in a greeting interaction in order to illustrate some of the concepts involved in structured interaction.

This description was based on films that mainly featured middle class professionals in the eastern US interacting. This, of course, suggests that at least some features of interactions as reported in the study apply only to this particular culture. While interactions vary in different contexts and cultures this study still provides a useful example of how interactions can be structured: however, it should be noted that interpretations of the purposes of various acts are preliminary and tentative.

Three major phases of the greeting interaction are identified: the distance salutation, the approach and the close salutation. In some cases the greeting ends with the distance salutation but in others it is followed by the approach and the close salutation: thus the structure of the greeting may vary with certain components being omitted in certain situations.

The distance salutation is so named because it is always followed by a phase in which the interactants move closer to one another. This act establishes that the interactants are now engaged in greeting one another. There are a number of different acts available to fill the slot: head tosses, head lowers, nodding and waving were all observed. Different kinds of salutation may have different functions: for instance, the nod is commonly observed as a 'salutation-in-passing' that is not followed by the other stages of greeting described here.

The approach occurs between the distance salutation and the close salutation and appears to contain two sub-phases. In the first of these behaviours associated with withdrawal occur, such as looking away from the other: in particular there is a sharp look-away immediately before the second sub-phase begins. This also tends to occur just before the point where utterance exchange becomes possible. In this case, then, there is both a kinesic and a proxemic indication of progression to the next sub-phase of the interaction. During the second phase of the approach vocalisation may occur. This phase is also associated with facial and other kinesic signals.

The close salutation occurs in its own location after the approach has been completed. Some form of ritual interaction is engaged in, three broad kinds of which are distinguished: close salutation without body contact, handshakes and kisses and embraces. When the close salutation is completed the interactants perform an S/O shift although this shift is sometimes very subtle and therefore difficult to detect. When this shift occurs the greeting is considered to be over.

4 Applications of Interactional Structure

In the interactive vision approach we propose applying interactional structure in two ways: to interpret the interactional behaviour of other agents and to generate appropriate behaviours for an interactive robot. The discussion here focuses on the visual modality: however, interaction is a multi-channel process (Birdwhistell, 1970) and the approach outlined here could be usefully extended by incorporating other modalities. Indeed it may well be

that visual behaviours can only be completely understood in association with behaviours that occur simultaneously in other modalities (Schefflen, 1973).

The proxemic factors outlined above seem reasonably easy to detect using a machine vision system. Similarly it is possible, though harder, to detect kinesic behaviours. Using this information we can make inferences about the type of interaction that is being observed in a scene. Once we have identified the interaction and its current phase then gesture becomes easier to identify as the search space may be reduced to behaviours associated with that phase. Equally, if we observe a gesture occurring within a given slot of a particular interaction then we can interpret it in terms of its meaning in that context: this is important as the meaning of behaviours will vary with context (Birdwhistell, 1970; Kendon, 1980).

Ideally such a system would be able to entertain multiple hypotheses regarding the interaction. This would allow it to deal with cases where an action incompatible with the current hypothesis occurs by switching to a hypothesis that is compatible with both the prior observations and the new ones. It would also be both top-down and bottom-up in nature: visual data can be interpreted as a specific action, leading to a particular assumption about the kind of action that is occurring which would in turn have implications for the interaction that is currently assumed to be taking place. This assumption would then be used to guide the interpretation of future input data (see figure 2).

There are, of course, a number of difficulties to be overcome. Firstly, neither kinesic nor proxemic behaviour is culture-independent so that a system based on the behaviours of one culture may not apply to other cultures. Equally, the appropriate interpretation of an action may vary with context (e.g. a cheery greeting may be appropriate in many circumstances, but not at a funeral).

A second difficulty is that people cannot be expected to separate their actions into simple discrete units. A system to analyse interactional behaviour needs to be capable of identifying multiple overlapping actions separately. It may also be naïve to assume that the meaning of a combination of actions is the same as the meanings of the same actions when they occur in isolation.

Thirdly, no exhaustive list of possible actions occurring in each phase of an interaction is available. The system needs to be capable of dealing with cases where an observed action fails to match any of those identified as belonging to the current phase. There are two solutions here: the first would be simply to expand the search to cover all actions that the system is capable of recognising. This may result in successful identification of the action but this would not be guaranteed: furthermore, interpretation of its meaning in the unusual context would remain a problem. The second solution is to attempt a deeper level of analysis.

Collett (1983) presents one example of such a deeper analysis. In this paper features of salutations of respect are described: such salutations tend to be asymmetrical and iconically submissive. An asymmetrical action is one in which one interactant carries out an action that the other does not, such as crouching down before the other. Iconic submissiveness may show itself in body lowering and in the generation of more activity than the other – again, the crouch is a good example of an iconically submissive interaction. By detecting features such as these it may be possible to interpret and classify previously unknown actions to at least some extent. This approach is similar in principle to that described in Brand and Essa (1995), where the idea of identifying the metaphorical meanings of gestures based on their physical nature is presented (e.g. a dismissive gesture may appear as though an imaginary object is being brushed away). Knowledge of the current phase may also be helpful in interpretation of novel actions: if we are in the close salutation phase of a greeting, for example, then we can assume that the action is some

form of close salutation as this is the only kind of action identified with this phase.

All of the above, however, presupposes some means of identifying an ‘action’: this is a very high-level concept and it is hard to see how a significant action could be distinguished from incidental movement. In certain cases, though, there is a very narrow frame within which an action must occur, such as in the case of close salutations: if no action is detected in this frame it is reasonable to assume that whatever movement did occur constituted some kind of close salutation. Sometimes, though, there may be little to no visual

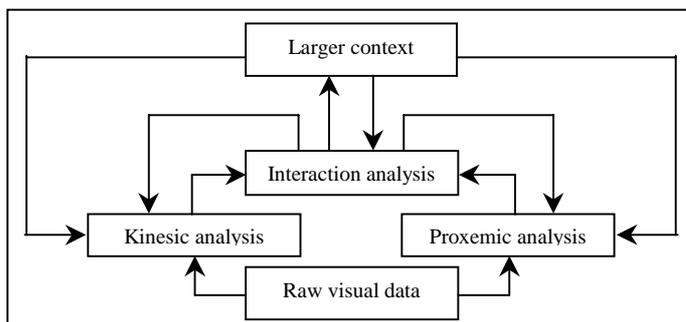


Figure 2: Low-level analysis leads to identification of higher-level features which, in turn, are used to guide future low-level analysis

behaviour. Here it would be helpful to have a system operating in multiple modalities: while there may be no visual behaviour in some close salutations there is still likely to be a verbal greeting.

The second way in which we propose that structured interaction may be used is action generation. By using knowledge of structured interaction a robot can generate the appropriate actions in response to an agent engaging in interaction with it, whether human or otherwise. Of course it will not necessarily be the case that the agent will replicate human behaviours: indeed, it may well be incapable of this (e.g. if it is not humanoid). However it can attempt to behave in a functionally equivalent way: for example the first sub-phase of the approach in greeting contains behaviours associated with withdrawal. Rather than duplicating human glance patterns or other appropriate behaviours the robot might simply approach in a tentative manner, e.g. by slowing down or repeatedly stopping and starting. At the very least, any robot that is capable of movement can employ proxemic behaviours.

However, there are again problems. While application of interactional structure in this manner may make interaction with robots more natural and efficient (as it does in interactions with humans), this will only work if humans are prepared to respond to the robots as they would to humans. Particularly in the case of non-humanoids it may be that interactions with pets would be a more appropriate model for robots of certain appearance and intelligence and in some cases it may be that entirely new interactional structures will need to be created. Still, Rintel and Pittam (1997) describe how humans have already applied some of the rules for one kind of interaction (face to face interactions) in a new domain (internet relay chat): it does seem, therefore, that humans can employ existing interactional structures in new situations.

There are also likely to be difficulties with attempts to use proxemic principles in non-humanoids. In humans proxemic behaviour is largely due to the desire for different kinds of sensory involvement at different distances. An artificial agent that simply moved to the American distance for close personal distance (Hall, 1966), for example, would not create the same visual impression as a human at the same distance unless it was a humanoid. Equally, this may not be the ideal distance for the kind of sensory information that the robot needs in order to interact effectively. While artificial agents can certainly engage in S/O shifts and use proxemic behaviour to indicate varying foci of attention, many aspects of proxemic behaviour cannot simply be mapped from human-human interactions to human-robot interactions. Again, human interactions with non-humans such as pets could prove a useful model here.

5 AuRoRA

This work is being carried out as a part of the AuRoRA project to develop a robot as a tool in the rehabilitation of autistic children (AuRoRA, 2000; Dautenhahn, 1999; Dautenhahn and Werry, 2000; Werry and Dautenhahn, 1999). One of the key deficits in autism is serious difficulty in dealing with social interactions: autistic children tend to favour structured and predictable environments and interactions and one of the goals of the AuRoRA project is that interaction with the more predictable robot should serve as a stepping stone to interaction with humans. Using a structured interactional framework should result in reasonably predictable behaviour as a distinct set of phases must be moved through: also, use of this framework should make the robot's interactive behaviour more similar to that of humans. The possibility of using different actions to fill the same slots also creates the potential for varying degrees of unpredictability in the robot's behaviour. As the robot will recognise and respond to nonverbal behaviour it should be possible to encourage the use of nonverbal expression in the children while its use of interactional structure may help to develop an understanding of such structures.

The robot being used in this project is an Applied AI Systems Labo-1, which is essentially a small flat-topped buggy (see figure 3). This limits the behaviours that the robot can employ: obviously it has a very limited capacity to perform kinesic behaviours, although the sensor array on the front can be made to turn from side to side. Most of the robot's nonverbal behaviours, though, will necessarily be proxemic in nature. At the present time the robot is only capable of sensing the children through fixed infrared sensors and a positional heat sensor. The robot is also capable of speech, greatly enhancing its interactive potential, but this paper is concerned with nonverbal behaviour.

The children are free to interact with the robot in an unconstrained manner: this poses a particular challenge for the development of a vision system as most systems that attempted detailed tracking of the human body



Figure 3: The AuRoRA robot with an autistic child: interactions between child and robot are highly unconstrained.

are able to assume that the human will remain at a given location and in the same approximate position (e.g. standing facing the camera and only moving the arms). Often only a part of the body is tracked (e.g. just the upper body) to reduce the complexity of models and thus save computation. These constraints clearly cannot be relied upon in this project: the child is engaging in full-body movement and will move around the room at will, shifting between standing and kneeling / crawling positions. These complexities are a part of the reason why this project will be largely dependent on proxemic rather than kinesic information. There is also the added complication of the presence of other moving objects in the scene: while many existing human tracking systems assume that only the person being tracked is moving, in this case both child and robot are moving. Additionally, two to three observers are usually present: these observers will also be moving within the scene observed by the vision system.

An obvious problem with attempting to apply the interactive vision approach with autistic individuals is that this approach is based on the nonverbal behaviour and interactional structures of psychologically 'normal' people. There is no guarantee that autistic individuals will behave in the same way: indeed, abnormal movement patterns have been associated with autism (AUTCOM, 2000; Teitelbaum et al, 1998) and autistic use and recognition of gestures and facial expression appears to be limited (Hobson, 1986a; Hobson, 1986b; Attwood, Frith and Hermelin, 1988). It seems likely that the interactive behaviour of the robot will at least need to be quite simple and it is possible that the use of this approach will not be particularly helpful in the AuRoRA project. However, even in this case the approach outlined here should still be useful more generally in the field of human-robot interaction.

Currently experiments are in preparation to test the usefulness of the concepts described in this paper. These experiments will focus on interactions between humans (both autistic and non-autistic) and the robot and will focus on proxemic features of behaviour such as relative position and orientation of interactants and use of S/O shifts. By explicitly observing these features of behaviour and interpreting them in terms of a model of a particular kind of interaction (e.g. greetings), the effectiveness of the vision system should be enhanced as outlined above. It is also intended that interactions between pairs of humans will be investigated: this will allow us to determine to what extent the social model can be independent of the types of agents involved in the interaction. We suspect that while the broad structure of the interaction will remain the same, humans will not interact with robots in exactly the same way as they would with other humans: as the robot neither looks nor behaves in a human manner humans are likely to treat them differently, as outlined above, although the work of Rintel and Pittam (1997) suggests that humans will nonetheless attempt to apply existing interactional structures in the new situation. A part of the challenge of designing behaviours for the robot will be getting it to behave in ways that humans will recognise as equivalent to a human behaviour in the current phase of the current interactional structure.

6 Conclusion

In conclusion, this paper has proposed an *interactive vision* approach to enhancing the abilities of robots to interact with human beings through the use of visual nonverbal signals that provide information about interactional structure. Such nonverbal signals can be used by robots (or, indeed, non-robotic artificial agents) in generating their own behaviour, in order to make them appear more believable as interaction partners and to allow them to engage in more effective interactions (as in Benford and Fahlén (1993)). These signals can also be used to interpret the interactional behaviour of others, whether interaction partners of the robot or other interacting agents. A number of these nonverbal signals can be identified by even quite simple vision systems as they are proxemic in nature, although the more complex kinesic signals should prove an increasingly useful part of this approach as machine vision technology develops. While there is as yet no experimental data to report, it is proposed that this is a promising approach to enhancing the social behaviour of artificial agents.

8 Acknowledgements

This project is supported by EPSRC GR/M62648 and conducted with the assistance of Radlett Lodge School of the National Autistic Society. The Labo-1 robot used in this research is donated by Applied AI Systems, Inc.

References

- Aggarwal, JK and Cai, Q, 1999, Human motion analysis: A review, *Computer Vision and Image Understanding* 73(3), pp 428-440
- Attwood, A, Frith, U and Hermelin, B, 1988, The understanding and use of interpersonal gestures by autistic and Down's syndrome children, *Journal of Autism and Developmental Disorders* 18(2), pp241-257
- AuRoRA, URL: <http://homepages.feis.herts.ac.uk/~comqkd/aurora.html>, last referenced 27th June 2000
- AUTCOM: Movement and relationship in autism, URL: <http://www.autcom.org/relationship.html>, last referenced 29th March 2000
- Azarbayejani, A, Wren, C and Pentland, A, 1996, Real-time 3-D tracking of the human body, Technical report 374, MIT Media Lab, Perceptual Computing Group
- Benford, S and Fahlén, L, 1993, A spatial model of interaction in large virtual environments, *Proceedings of the Third European Conference on Computer Supported Cooperative Work (ECSCW'93), Milano, Italy, September 1993*
- Birdwhistell, RL, 1970, *Kinesics and Context: Essays on Body-Motion Communication*, Penguin Books Ltd, Harmondsworth, Middlesex, UK
- Brand, M, 1997, Coupled hidden Markov models for modelling interacting processes, Technical report 405, MIT Media Lab, Perceptual Computing Section
- Crowley, JL, 1997, Vision for man-machine interaction, *Robotics and Autonomous Systems* 19, pp347-358
- Davis, JW and Bobick, AF, 1997, The representation and recognition of action using temporal templates, Technical report 402, MIT Media Lab, Perceptual Computing Group
- Brand, M and Essa, I, 1995, Causal analysis for visual gesture understanding, Technical report 327, MIT Media Lab, Perceptual Computing Group
- Breazeal, C and Fitzpatrick, P, 2000, That certain look: Social amplification of animate vision, To appear in *Proc. AAAI Fall Symposium "Socially Intelligent Agents - The Human in the Loop", 3-5 November, North Falmouth, MA, USA, Kerstin Dautenhahn (ed.)*
- Collett, P, 1983, Mossi salutations, *Semiotica*, 45, pp 191-248
- Dautenhahn, K, 1999, Robots as social actors: AuRoRA and the case of autism, *Proceedings Third Cognitive Technology Conference CT'99, August 1999, San Francisco*, pp 359-374
- Dautenhahn, K and Werry, I, 2000, Issues of robot-human interaction in the rehabilitation of children with autism, To appear in *Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior (SAB 2000), 11-15 September 2000, Paris, France*
- Ekman, P and Friesen, WV, 1969, The repertoire of nonverbal behavior: Categories, origins, usage and coding, *Semiotica* 1, pp49-98
- Essa, IA, 1999, Computers seeing people, *AI Magazine*, 20(1), pp 69-82
- Gavrila, DM and Davis, LS, 1995, Towards 3-D model-based tracking and recognition of human movement: A multi-view approach, *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, Zurich*
- Gill, SP, Kawamori, M, Katagiri, Y and Shimojima, A, 1999, Pragmatics of body moves, *Proceedings of the Third International Cognitive Technology Conference, San Francisco / Silicon Valley, USA*. Also available from <http://www.added.com.au/cogtech/CT99/Satinder.htm>, last referenced June 26th, 2000
- Hall, ET, 1966, *The Hidden Dimension: Man's Use of Space in Public and Private*, The Bodley Head Ltd, London, UK
- Hobson, RP, 1986a, The autistic child's appraisal of expressions of emotion, *Journal of Child Psychology and Psychiatry* 27(3), pp321-342
- Hobson, RP, 1986b, The autistic child's appraisal of expressions of emotion: A further study, *Journal of Child Psychology and Psychiatry* 27(5), pp671-680
- Johnson, N, Galata, A and Hogg, D, 1998, The acquisition and use of interaction behaviour models, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 1998*
- Ju, SX, Black, MJ and Yacoob, Y, 1996, Cardboard people: A parameterized model of articulated image motion, *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, IEEE*
- Kakadiaris, IA and Metaxas, D, 1998, Three-dimensional human body model acquisition from multiple views, *International Journal of Computer Vision*, 30(3), pp 191-218

- Kendon, A, 1980, Features of the structural analysis of human communicational behaviour, In W von Raffler-Engel (ed.), *Aspects of Nonverbal Communication*, Swets and Zeitlinger, Lisse, Netherlands, pp 29-43
- Kendon, A, 1990, *Conducting Interaction: Patterns of Behavior in Focused Encounters*, Cambridge University Press, Cambridge, UK
- Moore, DJ, Essa, IA and Hayes, MH, 1999, Exploiting human actions and object context for recognition tasks, Technical report GIT-GVU-99-11, Georgia Institute of Technology, Graphics, Visualisation and Usability Centre
- Oliver, N, Rosario, B and Pentland, A, 1999, *A Bayesian computer vision system for modelling human interactions*, Technical report 459, MIT Media Lab, Perceptual Computing Section
- Rintel, ES and Pittam, J, 1997, Strangers in a strange land: Interaction management on internet relay chat, *Human Communication Research* 23(4), pp 507-534
- Schank, RC and Abelson, R, 1977, *Scripts, plans, goals and understanding*, Lawrence Erlbaum Associates Inc, Hillsdale, NJ
- Schefflen, AE, 1973, *How Behavior Means*, Gordon and Breach Science Publishers Inc, New York, NY
- Seitz, SM and Dyer, CR, 1997, View-invariant analysis of cyclic motion, *International Journal of Computer Vision* 25(3), pp 231-251
- Teitelbaum, P, Teitelbaum, O, Nye, J, Fryman J and Maurer, RG, 1998, Movement analysis in infancy may be useful for early diagnosis of autism, *Proceedings of the National Academy of Sciences of the USA* 95, pp 13982-13987
- Werry, I and Dautenhahn, K, 1999, Applying mobile robot technology to the rehabilitation of autistic children, *Proceedings of the Symposium on Intelligent Robotic Systems (SIRS'99), Coimbra, Portugal*, pp 265-272
- Wren, CR and Pentland, AP, 1998, Dynamic models of human motion, Technical report 415, MIT Media Lab, Perceptual Computing Group