Behavioral Overlays for Non-Verbal Communication Expression on a Humanoid Robot

Andrew G. Brooks
Robotic Life Group
MIT Media Laboratory
77 Massachusetts Avenue, Cambridge, MA, USA
zoz@media.mit.edu

Ronald C. Arkin
Mobile Robot Laboratory and GVU Center
College of Computing, Georgia Institute of Technology
Atlanta, GA, USA
arkin@cc.gatech.edu

July 26, 2006

Abstract

This research details the application of non-verbal communication display behaviors to an autonomous humanoid robot, including the use of proxemics, which to date has been seldom explored in the field of human-robot interaction. In order to allow the robot to communicate information non-verbally while simultaneously fulfilling its existing instrumental behavior, a "behavioral overlay" model that encodes this data onto the robot's pre-existing motor expression is developed and presented. The state of the robot's system of internal emotions and motivational drives is used as the principal data source for non-verbal expression, but in order for the robot to display this information in a natural and nuanced fashion, an additional para-emotional framework has been developed to support the individuality of the robot's interpersonal relationships with humans and of the robot itself. An implementation on the Sony QRIO is described which overlays QRIO's existing EGO architecture and situated schema-based behaviors with a mechanism for communicating this framework through modalities that encompass posture, gesture and the management of interpersonal distance.

1 Introduction

When humans interact with other humans, they use a variety of implicit mechanisms to share information about their own state and the state of the interaction. Expressed over the channel of the physical body, these mechanisms are collectively known as non-verbal communication or "body language". It has not been proven that humans respond in precisely the same way to the body language of a humanoid robot as they do to that of a human. Nor have the specific requirements that the robot must meet in order to ensure such a response been empirically established. It has, however, been shown that humans will apply a social model to a sociable robot(Breazeal, C. 2003), and will in many cases approach interactions with electronic media holding a set of preconceived expectations based on their experiences of interactions with other humans(Reeves, B. and Nass, C. 1996). If these social equivalences extend to the interpretation of human-like body language displayed by a robot, it is likely that there will be corresponding benefits associated with enabling robots to successfully communicate in this fashion. For a robot such as the Sony QRIO (Figure 1) whose principal function is interaction with humans, we identify three such potential benefits.

First is the practical benefit of increasing the data bandwidth available for the "situational awareness" of the human, by transmitting more information without adding additional load to existing communication mechanisms. If it is assumed that it is beneficial for the human to be aware of the internal state of the robot, yet there are cases in which it is detrimental for the robot to interrupt other activities (e.g. dialog) in order to convey this information, an additional simultaneous data channel is called for. Non-verbal communication is an example of such a channel, and provided that the cues are implemented according to cultural norms and are convincingly expressible by the robot, adds the advantage of requiring no additional training for the human to interpret.

The second benefit is the forestalling of miscommunication within the expanded available data bandwidth. The problem raised if humans do indeed have automatic and unconscious expectations of receiving state information through bodily signals, is that certain states are represented by null signals, and humans interacting with a robot that does not communicate non-verbally (or which does so intermittently or ineffectively) may misconstrue lack of communication as deliberate communication of such a state. For example, failing to respond to personal verbal communications with attentive signals, such as eye contact, can communicate coldness or indifference. If a humanoid robot is equipped with those sorts of emotions, it is imperative that we try to ensure that such "false positive" communications are avoided to the fullest extent possible.

Figure 1: Sony QRIO, an autonomous humanoid robot designed for entertainment and interaction with humans, shown here in its standard posture.

The third potential benefit is an increased probability that humans will be able to form bonds with the robot that are analogous to those formed with other humans — for example, affection and trust. We believe that the development of such relations requires that the robot appear "natural" — that its actions can be seen as plausible in the context of the internal and external situations in which they occur. In other words, if a person collaborating with the robot can "at a glance" gain a perspective of not just what the robot is doing but why it is doing it, and what it is likely to do next, we think that he or she will be more likely to apply emotionally significant models to the robot. A principal theory concerning how humans come to be so skilled at modeling other minds is Simulation Theory, which states that humans model the motivations and goals of an observed agent by using their own cognitive structures to mentally simulate the situation of the observee (Davies, M. and Stone, T. 1995, Gordon, R. 1986, Heal, J. 2003). This suggests that it is likely that the more the observable behavior of the robot displays its internal state by referencing the behaviors the human has been conditioned to recognize — the more it "acts like" a human in its "display behaviors" — the more accurate the human's mental simulation of the robot can become.

With these benefits in mind as ultimate goals, we hereby report on activities towards the more immediate goal of realizing the display behaviors themselves, under three specific constraints. First, such displays should not restrict the successful execution of "instrumental behaviors", the tasks the robot is primarily required to perform. Second, the application of body language should be tightly controlled to avoid confusing the human — it must be expressed when appropriate, and suppressed when not. Third, non-verbal communication in humans is subtle and complex; the robot must similarly be able to use the technique to represent a rich meshing of emotions, motivations and memories. To satisfy these requirements, we have developed the concept of behavioral "overlays" for incorporating non-verbal communication displays into pre-existing robot behaviors. First, overlays provide a practical mechanism for modifying the robot's pre-existing activities "on-the-fly" with expressive information rather than requiring the design of specific new activities to incorporate it. Second, overlays permit the presence or absence of body language, and the degree to which it is expressed, to be controlled independent of the underlying activity. Third, the overlay system can be driven by an arbitrarily detailed model of these driving forces without forcing this model to be directly programmed into every underlying behavior, allowing even simple activities to become more nuanced and engaging.

A brief summary of the contributions of this research follows:

1. This work broadens and reappraises the use of bodily expressiveness in humanoid robots, particularly in the form of proxemics, which has hitherto been only minimally considered due to safety considerations and the relative scarcity of mobile humanoid platforms.

2. This work introduces the concept of a behavioral overlay for non-verbal communication that both encodes state information into the physical output of ordinary behaviors without requiring

modification to the behaviors themselves, and increases non-verbal bandwidth by injecting additional communicative behaviors in the absence of physical resource conflicts.

3. This paper further develops the behavioral communications overlay concept into a general model suitable for application to other robotic platforms and information sources.

4. In contrast with much prior work, the research described here provides more depth to the information that is communicated non-verbally, giving the robot the capability of presenting its internal state as interpreted via its own individuality and interpersonal memory rather than simply an instantaneous emotional snapshot.

5. Similarly, while expressive techniques such as facial feature poses are now frequently used in robots to communicate internal state and engage the human, this work makes progress towards the use of bodily expression in a goal-directed fashion.

6. Finally, this work presents a functioning implementation of a behavioral overlay system on a real robot, including the design of data structures to represent the robot's individual responsiveness to specific humans and to its own internal model.

2 Proxemics and Body Language

Behavioral researchers have comprehensively enumerated and categorized various forms of non-verbal communication in humans and animals. In considering non-verbal communication for a humanoid robot, we have primarily focused on the management of spatial relationships and personal space (proxemics) and on bodily postures and movements that convey meaning (kinesics). The latter class will be more loosely referred to as "body language", to underscore the fact that while many kinesic gestures can convey meaning in their own right, perhaps the majority of kinesic contributions to non-verbal communication occurs in a paralinguistic capacity, as an enhancement of concurrent verbal dialog(Dittmann, A. 1978). The use of this term is not intended, however, to imply that these postures and movements form a true language with discrete rules and grammars; but as Machotka and Spiegel point out, they convey coded messages that humans can interpret(Machotka, P. and Spiegel, J. 1982). Taken together, proxemics and body language can reflect some or all of the type of interaction, the relations between participants, the internal states of the participants and the state of the interaction.

2.1 Proxemics

Hall, pioneer of the field of proxemics, identified a number of factors that could be used to analyze the usage of interpersonal space in human-human interactions(Hall, E.T. 1966). State descriptors include the potential for the participants to touch, smell and feel the body heat of one another, and the visual appearance one another's face at a particular distance (focus, distortion, domination of visual field). The reactions of individuals to particular proxemic situations were documented according to various codes, monitoring aspects such as the amount and type of visual and physical contact, and whether or not the body posture of the subjects was encouraging ("sociopetal") or discouraging ("sociofugal") of such contact.

An informal classification was used to divide the continuous space of interpersonal distance into four general zones according to these state descriptors. In order of increasing distance, these are "Intimate", "Personal", "Socio-Consultive" and "Public". Human usage of these spaces in various relationships and situations has been observed and summarized(Weitz, S. 1974), and can be used to inform the construction of a robotic system that follows similar guidelines (subject to variations in cultural norms).

Spatial separation management therefore has practical effects in terms of the potential for sensing and physical contact, and emotional effects in terms of the comfort of the participants with a particular spatial arrangement. What constitutes an appropriate arrangement depends on the nature of the interaction (what kinds of sensing and contact are necessary, and what kinds of emotional states are desired for it), the relationship between the participants (an appropriate distance for a married couple may be different than that between business associates engaged in the same activity), and the current emotional states of the participants (the preceding factors being equal, the shape and size of an individual's ideal personal "envelope" can exhibit significant variation based on his or her feelings at the time, as shown in Figure 2).

To ensure that a robot displays appropriate usage of and respect for personal space, and to allow it to take actions to manipulate it in ways that the human

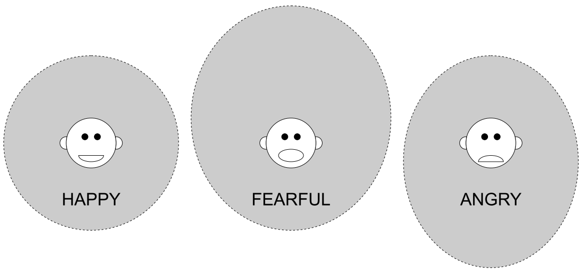

Figure 2: Illustration of how an individual's 'personal' space zone may vary in size and shape according to emotion. During fear, the space an individual considers his own might expand, with greater expansion occurring to his rear as he avoids potential threats that he cannot see. During anger, the space an individual considers her own might expand to a greater extent to her front as she directs her confrontational attention to known presences.

Proxemic Factor	Human	QRIO
Kinesthetic potential (arms only)	60–75cm	20cm
Kinesthetic potential (arms plus torso)	90–120cm	25–35cm
Minimum face recognition distance	< 5cm	20cm

Table 1: A comparison of select proxemic factors that differ between adult humans and QRIO (equipped with standard optics).

can understand and infer from them the underlying reasoning, requires consideration of all of these factors. In addition, the size of the robot (which is not limited to the range fixed by human biology) may have to be taken into account when considering the proxemics that a human might be likely to find natural or comfortable. See Table 1 for a comparison of several proxemic factors in the case of adult humans and QRIO, and Figure 3 for the general proxemic zones that were selected for QRIO.

There has been little exploration of the use of proxemics in human-robot interaction to date. The reasons for this are perhaps mostly pragmatic in nature. Robotic manipulators, including humanoid upper torsos, can be dangerous to humans and in most cases are not recommended for interactions within distances at which physical contact is possible. In addition, humanoid robots with full mobility are still relatively rare, and those with legged locomotion (complete humanoids) even more so, precluding such

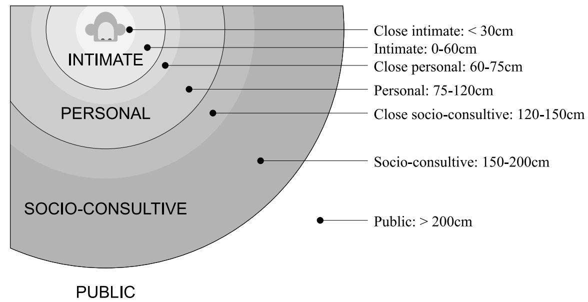

Figure 3: QRIO's proxemic zones in this implementation, selected as a balance between the zones for an adult human and those computed using QRIO's relevant proxemic factors. The demarcation distances between zones represent the midpoint of a fuzzy threshold function, rather than a 'hard' cutoff.

investigations. However, some related work exists.

The mobile robot 'Chaser' by Yamasaki and Anzai focused on one of the practical effects of interpersonal distance by attempting to situate itself at a distance from the human that was ideal for sensor operation, in this case the collection of speech audio(Yamasaki, N. and Anzai, Y. 1996). This work demonstrated that awareness of personal distance considerations could be acted upon to improve speech recognition performance.

In a similar vein, Kanda et al. performed a human-robot interaction field study in which the robot distinguished concurrently present humans as either participants or observers based on their proxemic distance(Kanda, T., Hirano, T., Eaton, D. and Ishiguro, H. 2004). A single fixed distance threshold was used for the classification, however, and it was left up to the human participants to maintain the appropriate proxemics.

Likhachev and Arkin explored the notion of "comfort zones" for a mobile robot, using attachment theory to inform an emotional model that related the robot's comfort to its spatial distance from an object of attachment(Likhachev, M. and Arkin, R.C. 2000). The results of this work showed that the robot's exploration behavior varied according to its level of comfort; while the published work did not deal with human-robot interaction directly, useful HRI scenarios could be envisaged for cases in which the object of attachment was a human.

More recently, there have been investigations into modifying the spatial behavior of non-humanoid mobile robots in order to make people feel more at ease with the robot. Smith investigated self-adaptation of the behavior of an interactive mobile robot, including its spatial separation from the human, based on its assessment of the human's comfort(Smith, C. 2005). In this case the goal was for the robot to automatically learn the personal space preferences of individual humans, rather than to use spatial separation as a general form of non-verbal communication. However some of the informal human subject experiments reported are instructive as to the value of taking proxemics into account in HRI, particularly the case in which misrecognition of a human discomfort response as a comfort response leads to a positive feedback loop that results in distressing behavior on the part of the robot.

Similar efforts have used considerations of humans' personal space to affect robot navigation. Nakauchi and Simmons developed a robot whose goal was to queue up to register for a conference along with human participants(Nakauchi, Y. and Simmons, R. 2000). The robot thus needed to determine how to move to positions that appropriately matched human queuing behavior. Althaus et al. describe a system developed to allow a robot to approach a group of people engaged in a discussion, enter the group by assuming a spatially appropriate position, and then leave and continue its navigation(Althaus, P., Ishiguro, H., Kanda, T., Miyashita, T. and Christensen, H.I. 2004). Christensen and Pacchierotti use proxemics to inform a control strategy for the avoidance behavior exhibited by a mobile robot when forced by a constraining passageway to navigate in close proximity to humans(Christensen, H.I. and Pacchierotti, E. 2005). Pacchierotti et al. then report positive results from a pilot user study, in which subjects preferred the condition in which the robot moved fastest and signaled the most deference to the humans (by moving out of the way earliest and keeping the greatest distance away), though the subjects were all familiar and comfortable with robots(Pacchierotti, E., Christensen, H.I. and Jensfelt, P. 2005). While informed by proxemics theory, these efforts focus on control techniques for applying social appropriateness to navigation activity, rather than using utilizing proxemic behavior as part of a non-verbal communication suite.

Another recent experiment, by te Boekhorst et al., gave some consideration to potential effects of the distance between children and a non-humanoid robot on the children's attention to a "pass the parcel" game(te Boekhorst, R., Walters, M., Koay, K.L., Dautenhahn, K. and Nehaniv, C. 2005). No significant effects were recorded, however the authors admit that the data analysis was complicated by violations of the assumptions underlying the statistical tests and therefore we believe these results should not be considered conclusive. Data from the same series of experiments was used to point out that the children's initial approach distances were socially appropriate according to human proxemics theory(Walters, M.L., Dautenhahn, K., Koay, K.L., Kaouri, C., te Boekhorst, R., Nehaniv, C.L., Werry, I. and Lee, D. 2005a). A follow-up experiment was conducted using the same non-humanoid robot to investigate the approach distances that adults preferred when interacting with the robot(Walters, M.L., Dautenhahn, K., te Boekhorst, R., Koay, K.L., Kaouri, C., Woods, S., Nehaniv, C.L., Lee, D. and Werry, I. 2005b). A majority, 60%, positioned themselves at a distance compatible with human proxemics theory, whereas the remainder, a significant minority of 40%, assumed positions significantly closer. While these results are generally encouraging in their empirical support of the validity of human-human proxemic theory to human-robot interactions, caution should be observed in extrapolating the results in either direction due to the non-humanoid appearance of the robot.

More recently than the initial submission of this paper, there has been interest shown in the use of proxemics and non-verbal communication by the robotic search and rescue community. In a conference poster Bethel and Murphy proposed a set of guidelines for affect expression by appearance-constrained (i.e. non-humanoid) rescue robots based on their proxemic zone with respect to a human(Bethel, C.L. and Murphy, R.R. 2006). The goal of this work was once again to improve the comfort level of humans through socially aware behavior.

While not involving robots per se, some virtual environment (VE) researchers have examined reactions to proxemic considerations between humans and humanoid characters within immersive VEs. Bailenson et al., for example, demonstrated that people exhibited similar personal spatial behavior towards virtual humans as they would towards real humans, and this effect was increased the more the virtual human was believed to be the avatar of a real human rather than an agent controlled by the computer(Bailenson, J.N., Blascovich, J., Beall, A.C. and Loomis, J.M. 2003). However, they also encountered the interesting result that subjects exhibited more pronounced avoidance behavior when proxemic boundaries were violated by an agent rather than an avatar; the authors theorize that this is due to the subjects attributing more rationality and awareness of social spatial behavior to a human-driven avatar than a computer-controlled agent, thus "trusting" that the avatar would not walk into their virtual bodies, whereas an agent might be more likely to do so. This may present a lesson for HRI designers: the precise fact that an autonomous robot is known to be under computer control may make socially communicative proxemic awareness (as opposed to simple collision avoidance) particularly important for robots intended to operate in close proximity with humans.

2.2 Body Language

Body language is the set of communicative body motions, or kinesic behaviors, including those that are a reflection of, or are intended to have an influence on, the proxemics of an interaction. Knapp identifies five basic categories:

1. Emblems, which have specific linguistic meaning and are what is most commonly meant by the term 'gestures';
2. Illustrators, which provide emphasis to concurrent speech;
3. Affect Displays, more commonly known as facial expressions and used to represent emotional states;
4. Regulators, which are used to influence conversational turn-taking; and
5. Adaptors, which are behavioral fragments that convey implicit information without being tied to dialog(Knapp, M. 1972).

Dittmann further categorizes body language into discrete and continuous (persistent) actions, with discrete actions further partitioned into categorical (always performed in essentially the same way) and non-categorical(Dittmann, A. 1978). Body posture itself is considered to be a kinesic behavior, inasmuch as motion is required to modify it, and because they can be modulated by attitude(Knapp, M. 1972).

The kinds of body language displays that can be realized on a particular robot of course depend on the mechanical design of the robot itself, and these categorizations of human body language are not necessarily of principal usefulness to HRI designers other than sometimes suggesting implementational details (e.g. the necessity of precisely aligning illustrators with spoken dialog). However, it is useful to examine the broad range of expression that is detailed within these classifications, in order to select those appearing to have the most utility for robotic applications. For example, classified within these taxonomies are bodily motions that can communicate explicit symbolic concepts, deictic spatial references (e.g. pointing), emotional states and desires, likes and dislikes, social status, engagement and boredom. As a result there has been significant ongoing robotics research overlapping with all of the areas thus referenced. For a comprehensive survey of socially interactive robotic research in general, incorporating many of these aspects, see (Fong, T., Nourbakhsh, I. and Dautenhahn, K. 2003).

The use of emblematic gestures for communication has been widely used on robotic platforms and in the field of animatronics; examples are the MIT Media Lab's 'Leonardo', an expressive humanoid which currently uses emblematic gesture as its only form of symbolic communication and also incorporates kinesic adaptors in the form of blended natural idle motions(Breazeal, C., Brooks, A.G., Gray, J., Hoffman, G., Kidd, C., Lee, H., Lieberman, J., Lockerd, A. and Chilongo, D. 2004), and Waseda University's WE-4RII 'emotion expression humanoid robot' which was also designed to adopt expressive body postures(Zecca, M., Roccella, S., Carrozza, M.C., Cappiello, G., Cabibihan, J.-J., Dario, P., Takanobu, H., Matsumoto, M., Miwa, H., Itoh, K. and Takanishi, A. 2004).

Comprehensive communicative gesture mechanisms have also been incorporated into animated humanoid conversational agents and VE avatars. Kopp and Wachsmuth used a hierarchical kinesic model to generate complex symbolic gestures from gesture phrases, later interleaving them tightly with concurrent speech(Kopp, S. and Wachsmuth, I. 2000, Kopp, S. and Wachsmuth, I. 2002). Guye-Vuilleme et al. provided collaborative VE users with the means to manually display a variety of non-verbal bodily expressions on their avatars using a fixed palette of potential actions(Guye-Vuilleme, A., Capin, T.K., Pandzic, I.S., Thalmann, N.M. and Thalmann, D. 1998).

Similarly, illustrators and regulators have been used to punctuate speech and control conversational turn-taking on interactive robots and animated characters. Aoyama and Shimomura implemented contingent head pose (such as nodding) and automatic filler insertion during speech interactions with Sony QRIO(Aoyama, K. and Shimomura,

H. 2005). Extensive work on body language for animated conversational agents has been performed at the MIT Media Lab, such as Thorisson's implementation of a multimodal dialog skill management system on an animated humanoid for face-to-face interactions (Thorisson, K.R. 1996) and Cassell and Vilhjalmsson's work on allowing human-controlled full-body avatars to exhibit communicative reactions to other avatars autonomously(Cassell, J. and Vilhjalmsson, H. 1999).

Robots that communicate using facial expression have also become the subject of much attention, too numerous to summarize here but beginning with well-known examples such as Kismet and the face robots developed by Hara(Breazeal, C. 2000, Hara, F., Akazawa, H. and Kobayashi, H. 2001). In addition to communication of emotional state, some of these robots have used affective facial expression with the aim of manipulating the human, either in terms of a desired emotional state as in the case of Kismet or in terms of increasing desired motivation to perform a collaborative task as in subsequent work on Leonardo(Brooks, A.G., Berlin, M., Gray, J. and Breazeal, C. 2005).

However much of this related work either focuses directly on social communication through body language as the central research topic rather than the interoperation of non-verbal communication with concurrent instrumental behavior, or on improvements to the interaction resulting from directly integrating non-verbal communication as part of the interaction design process. When emotional models are incorporated to control aspects such as affective display, they tend to be models designed to provide a "snapshot" of the robot's emotional state (for example represented by a number of discrete categories such as the Ekman model (Ekman, P. and Davidson, R.J. 1994)) suitable for immediate communication via the robot's facial actuators but with minimal reference to the context of the interaction. However recent research by Fridlund has strongly challenged the widely accepted notion that facial expressions are an unconscious and largely culturally invariant representation of internal emotional state, arguing instead that they are very deliberate communications that are heavily influenced by the context in which they are expressed(Fridlund, A. 1994). This is a contention that may be worth keeping in mind concerning robotic body language expression in general.

Given the capabilities of our robotic platform (Sony QRIO) and the relevance of the various types of body language to the interactions envisaged for QRIO, the following aspects were chosen for specific attention:

I. Proxemics and the management of interpersonal distance, including speed of locomotion;

II. Emblematic hand and arm gestures in support of the above;

III. The rotation of the torso during interaction, which in humans reflects the desire for interaction (facing more squarely represents a sociopetal stance, whereas displaying an angular offset is a sociofugal posture) — thus known as the "sociofugal/sociopetal axis";

IV. The posture of the arms, including continuous measures (arms akimbo, defensive raising of the arms, and rotation of the forearms which appear sociopetal when rotated outwards and sociofugal when rotated inwards) and discrete postural stances (e.g. arms folded);

V. Head pose and the maintenance of eye contact;

VI. Illustrators, both pre-existing (head nodding) and newly incorporated (attentive torso leaning).

See Figure 15 in Section 6 for examples of some of these poses as displayed by QRIO.

3 Behavioral Overlays

The concept of a behavioral overlay for robots can be described as a motor-level modifier that alters the resulting appearance of a particular output conformation (a motion, posture, or combination of both). The intention is to provide a simultaneous display of information, in this case non-verbal communication, through careful alteration of the motor system in such a way that the underlying behavioral activities of the robot may continue as normal. Just as the behavior schemas that make up the robot's behavioral repertoire typically need not know of the existence or method of operation of one another, behavioral overlays should be largely transparent to the behavior schema responsible for the current instrumental behavior at a given time. Preservation of this level of modularity simplifies the process of adding or learning new behaviors.

As a simplified example, consider the case of a simple 2-DOF output system: the tail of a robotic dog such as AIBO, which can be angled around horizontal and vertical axes. The tail is used extensively for non-verbal communication by dogs, particularly through various modes of tail wagging. Four examples of such communications via the tail from the AIBO motor primitives design specification are the following:

- *Tail Wagging Friendly*: Amplitude of wag large, height of tail low, speed of wag baseline slow but related to strength of emotion.
- *Tail Wagging Defensive*: Amplitude of wag small, height of tail high, speed of wag baseline fast but related to strength of emotion.
- *Submissive Posture*: In cases of low dominance/high submission, height of tail very low (between legs), no wagging motion.
- *"Imperious Walking"*: Simultaneous with locomotion in cases of high dominance/low submission; amplitude of wag small, height of tail high, speed of wag fast.

One method of implementing these communication modes would be to place them within each behavioral schema, designing these behaviors with the increased complexity of responding to the relevant emotional and instinct models directly. An alternative approach — behavioral overlays — is to allow simpler underlying behaviors to be externally modified to produce appropriate display activity. Consider the basic walking behavior $\mathbf{b}_0$ in which the dog's tail wags naturally left and right at height ϕ_0 with amplitude $\pm\theta_0$ and frequency ω_0 from default values for the walking step motion rather than the dog's emotional state. The motor state $\mathbf{M}_T$ of the tail under these conditions is thus given by:

$$\mathbf{M}_T(t) = \begin{bmatrix} \mathbf{b}_{0_x}(t) \\ \mathbf{b}_{0_y}(t) \end{bmatrix} = \begin{bmatrix} \theta_0 sin(\omega_0 t) \\ \phi_0 \end{bmatrix}$$

Now consider a behavioral overlay vector for the tail $\mathbf{o}_0 = [\alpha, \lambda, \delta]$ applied to the active behavior according to the mathematics of a multidimensional overlay coordination function Ω_T to produce the following overlaid tail motor state $\mathbf{M_T^+}$:

$$\mathbf{M}_T^+(t) = \Omega_T(\mathbf{o}_0, \mathbf{b}_0(t)) = \begin{bmatrix} \alpha\theta_0 sin(\lambda\omega_0 t) \\ \phi_0 + \delta \end{bmatrix}$$

For appropriate construction of $\mathbf{o}_0$, the overlay system is now able to produce imperious walking ($\alpha \ll 1$, $\lambda \gg 1$, $\delta \gg 0$) as well as other communicative walk styles not specifically predefined (e.g. "submissive walking": $\alpha = 0$, $\delta \ll 0$), without any modification of the underlying walking behavior. Moreover, the display output will continue to reflect as much as possible the parameters of the underlying activity (in this case the walking motion) in addition to the internal state used to generate the communications overlay (e.g. dominance/submission, emotion).

However in our AIBO example so far, the overlay system is only able to communicate the dog's internal state using the tail when it is already being moved by an existing behavior (in this case walking). It may therefore be necessary to add an additional special type of behavior whose function is to keep the overlay system supplied with motor input. This type of behavior is distinguished by two characteristics: a different stimulus set than normal behaviors (either more or less, including none at all); and its output is treated differently by the overlay system (which may at times choose to ignore it entirely). We refer to such behaviors as "idler" behaviors. In this case, consider idler behavior $\mathbf{b}_1$ which simply attempts to continuously wag the tail some amount in order to provide the overlay system with input to be accentuated or suppressed:

$$\mathbf{b}_1(t) = \begin{bmatrix} \theta_1 sin(\omega_1 t) \\ \phi_1 \end{bmatrix}$$

This behavior competes for action selection with the regular "environmental" behaviors as normal, and when active is overlaid by Ω_T in the same fashion. Thus the overlay system with the addition of one extremely basic behavior is able to achieve all of the four tail communications displays originally specified above, including variations in degree and combination, by appropriate selection of overlay components based on the robot's internal state. For active behavior $\mathbf{b}_i$ and overlay $\mathbf{o}_j$, the tail activity produced is:

$$\mathbf{M}_T^+(t) = \Omega_T(\mathbf{o}_j, \mathbf{b}_i(t))$$

However, the addition of specialized "idler" behaviors provides additional opportunities for manipulation of the robot's display activity, as these behaviors can be designed to be aware of and communicate with the overlay system — for example, to enable the triggering of emblematic gestures. If the robot's normal behaviors are subject at a given time to the stimulus vector $[\mathbf{S}]$, the idler behaviors can be thought of as responding to an expanded stimulus vector $[\mathbf{S}, \Psi]$

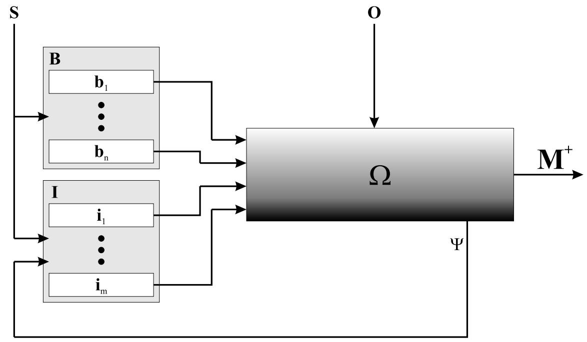

Figure 4: The behavioral overlay model, shown overlaying active environmental behaviors $\mathbf{b}_{1..n}$ and idler behaviors $\mathbf{i}_{1..m}$ after action selection has already been performed.

where Ψ is the vector of feedback stimuli from the overlay system. For instance, AIBO's tail idler behavior upon receiving a feedback stimulus ψ_p might interrupt its wagging action to trace out a predefined shape P with the tail tip:

$$\mathbf{b}_1(\psi_p, t) = \begin{bmatrix} P_x(t) \\ P_y(t) \end{bmatrix}$$

In general, then, let us say that for a collection of active (i.e. having passed action selection) environmental behaviors $\mathbf{B}$ and idler behaviors $\mathbf{I}$, and an overlay vector $\mathbf{O}$, the overlaid motor state $\mathbf{M}^+$ is given according to the model:

$$\mathbf{M}^+ = \Omega(\mathbf{O}, [\mathbf{B}(\mathbf{S}), \mathbf{I}([\mathbf{S}, \Psi])])$$

This model is represented graphically in Figure 4. In the idealized modular case, the environmental behaviors need neither communicate directly with nor even be aware of the existence of the behavioral overlay system. For practical purposes, however, a coarse level of influence by the environmental behaviors on the overlay system is required, because a complete determination a priori of whether or not motor modification will interfere with a particular activity is very difficult to make. This level of influence has been accounted for in the model, and the input to Ω from $\mathbf{B}$ and $\mathbf{I}$ as shown incorporates any necessary communication above and beyond the motor commands themselves.

Behavioral overlays as implemented in the research described in this paper include such facilities for schemas to communicate with the overlay system when necessary. Schemas may, if desired, protect themselves against modification in cases in which interference is likely to cause failure of the behavior (such as a task, like fine manipulation, that would similarly require intense concentration and suppression of the non-verbal communication channel when performed by a human). Care should, of course, be taken to use this facility sparingly, in order to avoid the inadvertent sending of non-verbal null signals during activities that should not require such concentration, or for which the consequences of failure are not excessively undesirable.

Schemas may also make recommendations to the overlay system that assist it in setting the envelope of potential overlays (such as reporting the characteristic proxemics of an interaction; e.g. whether a speaking behavior takes place in the context of an intimate conversation or a public address). In general, however, knowledge of and communication with the overlay system is not a requirement for execution of a behavior. As QRIO's intentional mechanism is refined to better facilitate high-level behavioral control, the intention system may also communicate with the behavioral overlay system directly, preserving the modularity of the individual behaviors.

Related work of most relevance to this concept is the general body of research related to motion parameterization. The essential purpose of this class of techniques is to describe bodily motions in terms of parameters other than their basic joint-angle time series. Ideally, the new parameters should capture essential qualities of the motion (such as its overall appearance) in such a way that these qualities can be predictably modified or held constant by modifying or holding constant the appropriate parameters. This has advantages both for motion generation (classes of motions can be represented more compactly as parameter ranges rather than clusters of individual exemplars) and motion recognition (novel motions can be matched to known examples by comparison of the parameter values).

Approaches to this technique have been applied to motion generation for animated characters can differ in their principal focus. One philosophy involves creating comprehensive sets of basic motion templates that can then be used to fashion more complex motions by blending and modifying them with a smaller set of basic parameters, such as duration, amplitude and direction; this type of approach was used under the control of a scripting language to add real-time gestural activity to the animated character OLGA(Beskow, J. and McGlashan, S. 1997). At the other extreme, the Badler research group at the University of Pennsylvania argued that truly lifelike motion requires the use of a large number of parameters concerned with effort and shaping, creating the animation model EMOTE based on Laban Movement Analysis; however the model is non-emotional and does not address autonomous action generation(Chi, D., Costa, M., Zhao, L. and Badler, N. 2000).

One of the most well known examples of motion parameterization that has inspired extensive attention in both the animation and robotics communities is the technique of "verbs and adverbs" proposed by Rose et al.(Rose, C., Cohen, M. and Bodenheimer, B. 1998). In this method verbs are specific base actions and adverbs are collections of parameters that modify the verbs to produce functionally similar motor outputs that vary according to the specific qualities the adverbs were designed to affect.

This technique allows, for example, an animator to generate a continuous range of emotional expressions of a particular action, from say 'excited waving' to 'forlorn waving', without having to manually create every specific example separately; instead, a single base 'waving' motion would be created, and then parameter ranges that described the variation from 'excited' to 'forlorn' provide the means of automatically situating an example somewhere on that continuum. While the essential approach is general, it is typically applied at the level of individual actions rather than overall behavioral output due to the difficulty of specifying a suitable parameterization of all possible motion.

Similarly, the technique of "morphable models" proposed by Giese and Poggio describes motion expressions in terms of pattern manifolds inside which plausible-looking motions can be synthesized and decomposed with linear parameter coefficients, according to the principles of linear superposition(Giese, M.A. and Poggio, T. 2000). In their original example, locomotion gaits such as 'walking' and 'marching' were used as exemplars to define the parameter space, and from this the parameters could be re-weighted in order to synthesize new gaits such as 'limping'. Furthermore, an observed gait could then be classified against the training examples using least-squares estimation in order to estimate its relationship to the known walking styles.

Not surprisingly, motion parameterization techniques such as the above have been shown significant interest by the segment of the robotics community concerned with robot programming by demonstration. Motion parameterization holds promise for the central problem that this approach attempts to solve: extrapolation from a discrete (and ideally small) set of demonstrated examples to a continuous task competency envelope; i.e., knowing what to vary to turn a known spatio-temporal sequence into one that is functionally equivalent but better represents current circumstances that were not prevailing during the original demonstration. The robotics literature in this area, even just concerning humanoid robots, is too broad to summarize, but see (Peters, R.A.II, Campbell, C.C., Bluethmann, W.J. and Huber, E. 2003) for a representative example that uses a verbs-adverbs approach for a learned grasping task, and illustrates the depth of the problem.

Fortunately, the problem that behavioral overlays seeks to address is somewhat simpler. In the first place, the task at hand is not to transform a known action that would be unsuccessful if executed under the current circumstances into one that now achieves a successful result; rather, it is to make modifications to certain bodily postures during known successful actions to a degree that communicates information without causing those successful actions to become unsuccessful. This distinction confers with it the luxury of being able to choose many parameters in advance according to well-described human posture taxonomies such as those referred to in Section 2, allowing algorithmic attention to thus be concentrated on the appropriate quantity and combination of their application. Furthermore, such a situation, in which the outcome even of doing nothing at all is at least the success of the underlying behavior, has the added advantage that unused bodily resources can be employed in overlay service with a reasonable level of assuredness that they will not cause the behavior to fail.

Secondly, the motion classification task, where it exists at all, is not the complex problem of parameterizing monolithic observed output motion-posture combinations, but simply to attempt to ensure that such conformations, when classified by the human observer's built-in parameterization function, will be classified correctly. In a sense, behavioral overlays start with motor descriptions that have already been parameterized — into the instrumental behavior itself and the overlay information — and the task of the overlay function is to maintain and apply this parameterization in such a fashion that the output remains effective in substance and natural in appearance, a far less ambiguous situation than the reverse case of attempting to separate unconstrained natural behavior into its instrumental and non-instrumental aspects (e.g. 'style' from 'content').

In light of the above, and since behavioral overlays are designed with the intention of affecting all of the robot's behavioral conformations, not just ones that have been developed or exhibited at the time of parameter estimation, the implementation of behavioral overlays described here has focused on two main areas. First, the development of general rules concerning the desired display behaviors and available bodily resources identified in Section 2 that can be applied to a wide range of the robot's activities. And second, the development of a maintenance and releasing mechanism that maps these rules to the space of emotional and other internal information that the robot will use them to express. Details of the internal data representations that provide the robot with the contextual information necessary to make these connections are given in Section 4, and implementation details of the overlay system itself are provided in Section 5.

4 Relationships and Attitudes

An important driving force behind proxemics and body language is the internal state of the individual. QRIO's standard EGO Architecture contains a system of internally-maintained emotions and state variables, and some of these are applicable to nonverbal communication. A brief review follows; please refer to Figure 5 for a block diagram of the unmodified EGO Architecture.

QRIO's emotional model (EM) contains six emotions (ANGER, DISGUST, FEAR, JOY, SADNESS, SURPRISE) plus NEUTRAL, along the lines of the Ekman proposal(Ekman, P. and Davidson, R.J. 1994). Currently QRIO is only able to experience one emotion from this list at a given time, though the non-verbal communication overlay system has been designed in anticipation of the potential for this to change. Emotional levels are represented by continuous-valued variables.

QRIO's internal state model (ISM) is also a system of continuous-valued variables, that are maintained within a certain range by a homeostasis mechanism. Example variables include FATIGUE, INFORMATION, VITALITY and INTERACTION. A low level of a particular state variable can be used to drive QRIO to seek objects or activities that can be expected to increase the level, and vice versa.

For more details of the QRIO emotionally grounded architecture beyond the above summary, see (Sawada, T., Takagi, T., Hoshino, Y. and Fujita, M. 2004). The remainder of this section details additions to the architecture that have been developed specifically to support non-verbal communication overlays. In the pre-existing EGO architecture, QRIO has been designed to behave differently with different individual humans by changing its EM and ISM values in response to facial identification of each human. However, these changes are instantaneous upon recognition of the human; the lack of additional factors distinguishing QRIO's feelings about these specific individual humans limits the amount of variation and naturalness that can be expressed in the output behaviors.

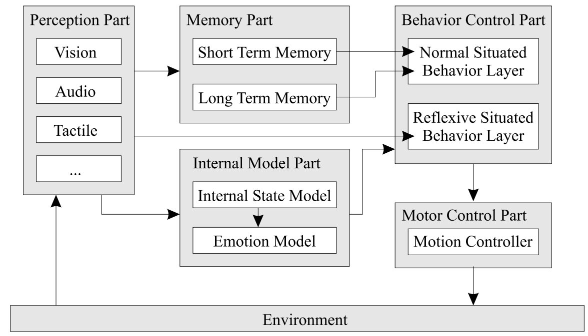

Figure 5: Functional units and interconnections in QRIO's standard EGO Architecture.

In order for QRIO to respond to individual humans with meaningful proxemic and body language displays during personal interactions, QRIO requires a mechanism for preserving the differences between these individual partners — what we might generally refer to as a relationship. QRIO does have a long-term memory (LTM) feature; an associative memory, it is used to remember connections between people and objects in predefined contexts, such as the name of a person's favorite food. To support emotional relationships, which can then be used to influence non-verbal communication display, a data structure to extend this system has been developed.

Each human with whom QRIO is familiar is represented by a single 'Relationship' structure, with several internal variables. A diagram of the structure can be seen in Figure 6. The set of variables chosen for this structure have generally been selected for the practical purpose of supporting non-verbal communication rather than to mirror a particular theoretical model of interpersonal relationships, with exceptions

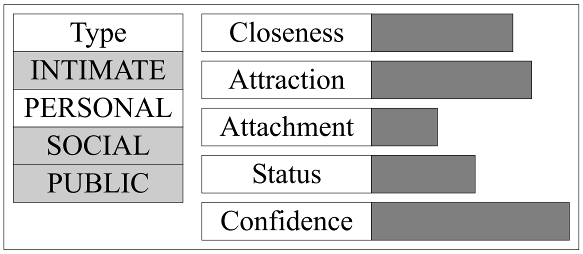

Figure 6: An example instantiation of a Relationship structure, showing arbitrarily selected values for the single discrete enumerated variable and each of the five continuous variables.

noted below.

The discrete-valued 'Type' field represents the general nature of QRIO's relationship with this individual; it provides the global context within which the other variables are locally interpreted. Because a primary use of this structure will be the management of personal space, it has been based on delineations that match the principal proxemic zones set out by Weitz(Weitz, S. 1974). An INTIMATE relationship signifies a particularly close friendly or familial bond in which touch is accepted. A PERSONAL relationship represents most friendly relationships. A SOCIAL relationship includes most acquaintances, such as the relationship between fellow company employees. And a PUBLIC relationship is one in which QRIO may be familiar with the identity of the human but little or no social contact has occurred. It is envisaged that this value will not change frequently, but it could be learned or adapted over time (for example, a PUBLIC relationship becoming SOCIAL after repeated social contact).

All other Relationship variables are continuous-valued quantities, bounded and normalized, with some capable of negative values if a reaction similar in intensity but opposite in nature is semantically meaningful. The 'Closeness' field represents emotional closeness and could also be thought of as familiarity or even trust. The 'Attraction' field represents QRIO's desire for emotional closeness with the human. The 'Attachment' field, based on Bowlby's theory of attachment behavior(Bowlby, J. 1969), is a variable with direct proxemic consequences and represents whether or not the human is an attachment object for QRIO, and if so to what degree. The 'Status' field represents the relative sense of superiority or inferiority QRIO enjoys in the relationship, allowing the structure to represent formally hierarchical relationships in addition to informal friendships and acquaintances. Finally, the Relationship structure has a 'Confidence' field which represents QRIO's assessment of how accurately the other continuous variables in the structure might represent the actual relationship; this provides a mechanism for allowing QRIO's reactions to a person to exhibit differing amounts of variability as their relationship progresses, perhaps tending to settle as QRIO gets to know them better.

In a similar vein, individual humans can exhibit markedly different output behavior under circumstances in which particular aspects of their internal states could be said to be essentially equivalent; their personalities affect the way in which their emotions and desires are interpreted and expressed. There is evidence to suggest that there may be positive outcomes to endowing humanoid robots with perceptible individual differences in the way in which they react to their internal signals and their relationships with people. Studies in social psychology and communication have repeatedly shown that people prefer to interact with people sharing similar attitudes and personalities (e.g. (Blankenship, V., Hnat, S.M., Hess, T.G. and Brown, D.R. 2004, Byrne, D. and Griffit, W. 1969)) — such behavior is known as "similarity attraction". A contrary case is made for "complementary attraction", in which people seek interactions with other people having different but complementary attitudes that have the effect of balancing their own personalities(Kiesler, D.J. 1983, Orford, J. 1986).

These theories have been carried over into the sphere of interactions involving non-human participants. In the product design literature, Jordan discusses the "pleasurability" of products; two of the categories in which products have the potential to satisfy their users are "socio-pleasure", relating to inter-personal relationships, and "ideo-pleasure", relating to shared values(Jordan, P.W. 2000). And a number of human-computer interaction studies have demonstrated that humans respond to computers as social agents with personalities, with similarity attraction being the norm (e.g. (Nass, C. and Lee, K.M. 2001)). Yan et al. performed experiments in which AIBO robotic dogs were programmed to simulate fixed traits of introversion or extroversion, and showed that subjects were able to correctly recognize the expressed trait; however, in this case the preferences observed indicated complementary attraction, with the authors postulating the embodiment of the robot itself as a potential factor in the reversal(Yan, C., Peng, W., Lee, K.M. and Jin, S. 2004).

As a result, if a humanoid robot can best be thought as a product which seeks to attract and ultimately fulfil the desires of human users, it might be reasonable to predict that humans will be more attracted to and satisfied by robots which appear to match their own personalities and social responses. On the other hand, if a humanoid robot is instead best described as a human-like embodied agent that is already perceived as complementary to humans as a result of its differences in embodiment, it might alternatively be predicted that humans will tend to be more attracted to robots having personalities that are perceptibly different from their own. In either case, a robot possessing the means to hold such attitudes would be likely to have a general advantage in attaining acceptance from humans, and ultimately the choice of the precise nature of an individual robot could be left up to its human counterpart.

To allow QRIO to exhibit individualistic (and thus hopefully more interesting) non-verbal behavior, a system that interprets or filters certain aspects of QRIO's internal model was required. At present this system is intended only to relate directly to the robot's proxemic and body language overlays; no attempt has been made to give the robot anything approaching a complete 'personality'. As such, the data structure representing these individual traits is instead entitled an 'Attitude'. Each QRIO has one such structure; it is envisaged to have a long-to-medium-term effect, in that it is reasonable for the structure to be pre-set and not to change thereafter, but it also might be considered desirable to have the robot be able to change its nature somewhat over the course of a particularly long term interaction (even a lifetime), much as human attitudes sometimes mellow or become more extreme over time. Brief instantiation of temporary replacement Attitude (and Relationship) structures would also provide a potential mechanism for QRIO to expand its entertainment repertoire with 'acting' ability, simply by using its normal mechanisms to respond to interactions as though they featured different participants.

The Attitude structure consists of six continuous-valued, normalized variables in three opposing pairs, as illustrated in Figure 7. Opposing pairs are directly related in that one quantity can be computed from the other; although only three variables are therefore computationally necessary, they are specified in this way to be more intuitively grounded from the point of view of the programmer or behavior designer.

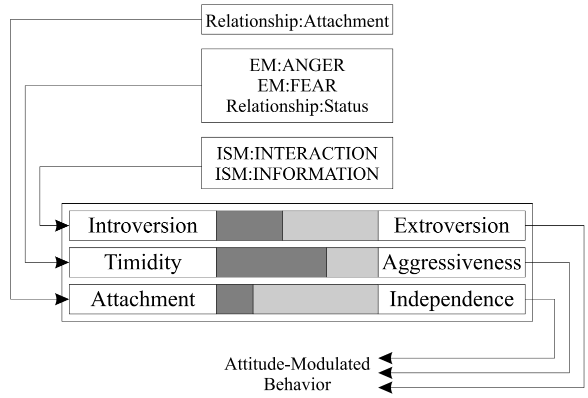

Figure 7: An arbitrary instantiation of an Attitude structure, showing the three opposing pairs of continuous-valued variables, and the ISM, EM and Relationship variables that each interprets.

The 'Extroversion' and 'Introversion' fields are adapted directly from the Extroversion dimension of the Five-Factor Model (FFM) of personality(McCrae, R.R. and Costa, P.T. 1996), and as detailed above were used successfully in experiments with AIBO. In the case of QRIO, these values are used to affect the expression of body language and to interpret internal state desires. High values of Extroversion encourage more overt body language, whereas high Introversion results in more subtle bodily expression. Extroversion increases the effect of the ISM variable INTERACTION and decreases the effect of INFORMATION, whereas Introversion does the reverse.

The 'Aggressiveness' and 'Timidity' fields are more loosely adapted from the FFM — they can be thought of as somewhat similar to a hybrid of Agreeableness and Conscientiousness, though the exact nature is more closely tailored to the specific emotions and non-verbal display requirements of QRIO. Aggressiveness increases the effect of the EM variable ANGER and decreases the effect of FEAR, while Timidity accentuates FEAR and attenuates ANGER. High Timidity makes submissive postures more probable, while high Aggressiveness raises the likelihood of dominant postures and may negate the effect of Relationship Status.

Finally, the 'Attachment' and 'Independence' fields depart from the FFM and return to Bowlby; their only effect is proxemic, as an interpreter for the value of Relationship Attachment. While any human can represent an attachment relationship with the robot, robots with different attitudes should be expected to respond to such relationships in different ways. Relationship Attachment is intended to have the effect of compressing the distance from the human that the robot is willing to stray; the robot's own Independence or Attachment could be used for example to alter the fall-off probabilities at the extremes of this range, or to change the 'sortie' behavior of the robot to bias it to make briefer forays away from its attachment object.

The scalar values within the Relationship and Attitude structures provide the raw material for affecting the robot's non-verbal communication (and potentially many other behaviors). These have been carefully selected in order to be able to drive the output overlays in which we are interested. However, there are many possible ways in which this material could then be interpreted and mathematically converted into the overlay signals themselves. We do not wish to argue for one particular numerical algorithm over another, because that would amount to claiming that we have quantitative answers to questions such as "how often should a robot which is 90% introverted and 65% timid lean away from a person to whom it is only 15% attracted?". We do not make such claims. Instead, we will illustrate the interpretation of these structures through two examples of general data usage models that contrast the variability of overlay generation available to a standard QRIO versus that available to one equipped with Relationships and Attitudes.

Sociofugal/sociopetal axis: We use a scalar value for the sociofugal/sociopetal axis, S, from 0.0 (the torso facing straight ahead) to 1.0 (maximum off-axis torso rotation). Since this represents QRIO's unwillingness to interact, a QRIO equipped with non-verbal communication skills but neither Relationships nor Attitudes might generate this axis based on a (possibly non-linear) function s of its ISM variable INTERACTION, I_{INT}, and its EM variable ANGER, E_{ANG}:

$$S = s(I_{INT}, E_{ANG})$$

The QRIO equipped with Relationships and Attitudes is able to apply more data towards this computation. The value of the robot's Introversion, A_{Int}, can decrease the effect of INTERACTION, resulting in inhibition of display of the sociofugal axis according to some combining function f. Conversely, the value of the robot's Aggression, A_{Agg}, can increase the effect of ANGER, enhancing the sociofugal result according to the combining function g. Furthermore, the robot may choose to take into account the relative Status, R_{Sta}, of the person with whom it is interacting, in order to politely suppress a negative display. The improved sociofugal axis generation function s^+ is thus given by:

$$S' = s^+(f(I_{INT}, A_{Int}), g(E_{ANG}, A_{Agg}), R_{Sta})$$

Proxemic distance: Even if the set of appropriate proxemic zones $\mathbf{P}$ for the type of interaction is specified by the interaction behavior itself, QRIO will need to decide on an actual scalar distance within those zones, D, to stand from the human. A standard QRIO might thus use its instantaneous EM variable FEAR, E_{FEA}, to influence whether it was prepared to choose a close value or to instead act more warily and stand back, according to another possibly non-linear function d:

$$D = d(\mathbf{P}, E_{FEA})$$

The enhanced QRIO, on the other hand, is able to make much more nuanced selections of proxemic distance. The proxemic Type field of the Relationship, R_{Typ}, allows the robot to select the most appropriate single zone from the options given in $\mathbf{P}$. The value of the robot's Timidity, A_{Tim}, increases the effect of FEAR, altering the extent to which the robot displays wariness according to a combining function u. The Closeness of the robot's relationship to the human, R_{Clo}, can also be communicated by altering the distance it keeps between them, as can its Attraction for the human, R_{Attr}. If the robot has an Attachment relationship with the human, its strength R_{Atta} can be expressed by enforcing an upper bound on D, modulated by the robot's Independence A_{Ind} according to the combining function v. The improved ideal proxemic distance generation function d^+ is thus given by:

$$D' = d^+\begin{pmatrix} \mathbf{P}, R_{Typ}, u(E_{FEA}, A_{Tim}), \\ R_{Clo}, R_{Attr}, v(R_{Atta}, A_{Ind}) \end{pmatrix}$$

Clearly, the addition of Relationships and Attitudes offer increased scope for variability of non-verbal communications output. More importantly, however, they provide a rich, socially grounded framework for that variability, allowing straightforward implementations to be developed that non-verbally communicate the information in a way that varies predictably with the broader social context of the interaction.

5 System Architecture and Implementation

The QRIO behavioral system, termed the EGO Architecture, is a distributed object-based software architecture based on the OPEN-R modular environment originally developed for AIBO. Objects in EGO are in general associated with particular functions. There are two basic memory objects: Short-Term Memory (STM) which processes perceptual information and makes it available in processed form at a rate of 2 Hz, and Long-Term Memory (LTM) which associates information (such as face recognition results) with known individual humans. The Internal Model (IM) object manages the variables of the ISM and EM. The Motion Controller (MC) receives and executes motion commands, returning a result to the requesting module.

QRIO's actual behaviors are executed in up to three Situated Behavior Layer (SBL) objects: the Normal SBL (N-SBL) manages behaviors that execute at the 2 Hz STM update rate (homeostatic behaviors); the Reflexive SBL (R-SBL) manages behaviors that require responses faster than the N-SBL can provide, and therefore operates at a significantly higher update frequency (behaviors requiring perceptual information must communicate with perceptual modules directly rather than STM); and deliberative behavior can be realized in the Deliberative SBL (D-SBL).

Within each SBL, behaviors are organized in a tree-structured network of schemas; schemas perform minimal communication between one another, and compete for activation according to a winner-take-all mechanism based on the resource requirements of individual schemas. For more information about the EGO Architecture and the SBL system of behavior control, please see (Fujita, M., Kuroki, Y., Ishida, T. and Doi, T. 2003).

5.1 NVC Object

Because behavioral overlays must be able to be applied to all behavior schemas, and because schemas are intended to perform minimal data sharing (so that schema trees can be easily constructed from individual schemas without having to be aware of the overall tree structure), it is not possible or desirable to completely implement an appropriate overlay system within the SBLs themselves. Instead, to implement behavioral overlays an additional, independent

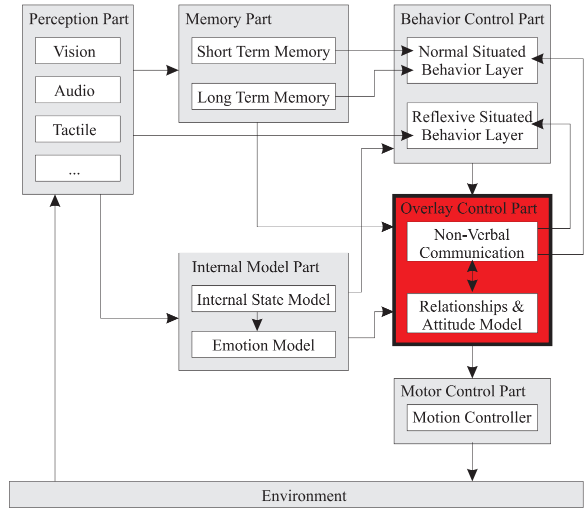

Figure 8: Interdependence diagram of the EGO Architecture with NVC.

Non-Verbal Communication (NVC) object was added to the EGO Architecture.

The NVC object has data connections with several of the other EGO objects, but its most central function as a motor-level overlay object is to intercept and modify motor commands as they are sent to the MC object. Addition of the NVC object therefore involves reconnecting the MC output of the various SBL objects to the NVC, and then connecting the NVC object to the MC. MC responses are likewise routed through the NVC object and then back to the SBLs.

In addition to the motor connection, the NVC object maintains a connection to the IM output (for receiving ISM and EM updates, which can also be used as a 2 Hz interrupt timer), the STM target update (for acquiring information about the location of humans, used in proxemic computations) and a custom message channel to the N-SBL and R-SBL (for receiving special information about the interaction, and sending trigger messages for particular gestures and postures). See Figure 8 for a graphical overview of the EGO Architecture with NVC. In addition to the connections shown, the NVC object manages the behavioral overlay values themselves with reference to the Relationship and Attitude structures; Figure 9 has an overview of the internal workings of NVC.

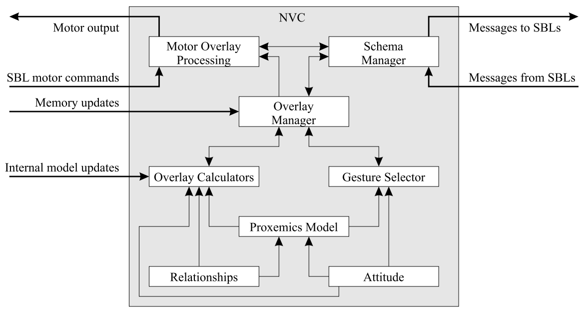

Figure 9: The non-verbal communication (NVC) module's conceptual internal structure and data interconnections with other EGO modules.

5.2 Overlays, Resources and Timing

The data representation for the behavioral overlays themselves is basic yet flexible. They are divided according to the major resource types Head, Arms, Trunk and Legs. For each joint (or walking parameter value, in the case of the legs) within a resource type, the overlay maintains a value and a flag that allows the value to be interpreted as either an absolute bias (to allow constant changes to the postural conformation) or a relative gain (to accentuate or attenuate incoming motions). In addition, each overlay category contains a time parameter for altering the speed of motion of the resource, which can also be flagged as a bias or a gain. Finally, the legs overlay contains an additional egocentric position parameter that can be used to modify the destination of the robot in the case of walking commands.

Motion commands that are routed through the NVC object consist of up to two parts: a command body, and an option parameter set. Motions that are parameterized (i.e., that have an option part) can be modified directly by the NVC object according to the current values of the overlays that the NVC object is storing. Such types of motions include direct positioning of the head, trunk and arm with explicit joint angle commands; general purpose motions that have been designed with reuse in mind, such as nodding (the parameter specifying the depth of nod); and commands that are intended to subsequently run in direct communication with perceptual systems with the SBL excluded from the decision loop, such as head tracking. Unfortunately due to the design of the MC system, unparameterized motion commands (i.e., those with just a body) cannot be altered before reaching the MC object; but they can be ignored or replaced with any other single parameterized or unparameterized motion command having the same actuator resource requirements (this possibility is not yet taken advantage of in the current implementation).

Resource management in the EGO Architecture is coarse grained and fixed; it is used for managing schema activation in the SBLs as well as just for presenting direct motion command conflicts. The resource categories in EGO are the Head, Trunk, Right Arm, Left Arm and Legs. Thus a body language motion command that wanted only to adjust the sociofugal/sociopetal axis (trunk rotate), for example, would nevertheless be forced to take control of the entire trunk, potentially blocking an instrumental behavior from executing. The NVC object implements a somewhat finer-grained resource manager by virtue of the overlay system. By being able to choose to modify the commands of instrumental behaviors directly, or not to do so, the effect is as if the resource were managed at the level of individual joints.

This raises, however, the problem of the case of time steps at which environmental behaviors do not send motion commands yet it is desired to modify the robot's body language; this situation is common. The NVC object skirts this problem by relying on a network of fast-activating idle schemas residing in the N-SBL. Each idle schema is responsible for a single MC resource. On each time step, if no instrumental behavior claims a given resource, the appropriate idle schema sends a null command to the NVC object for potential overlaying. If an overlay is desired, the null command is modified into a genuine action and passed on to the MC; otherwise it is discarded. Since commands from idle and other special overlay schemas are thus treated differently by the NVC object than those from behavioral schemas, the NVC object must keep track of which schemas are which; this is accomplished by a handshaking procedure that occurs at startup time, illustrated graphically in Figure 10. Normal operation of the idle schemas is illustrated in Figure 11.

For practical considerations within the EGO Architecture, actions involved in non-verbal communication can be divided into four categories according to two classification axes. First is the timing requirements of the action. Some actions, such as postural shifts, are not precisely timed and are appropriately suited to the 2 Hz update rate of the N-SBL. Others, however, are highly contingent with the activity, such as nodding during dialog, and must reside in the R-SBL. Second is the resource requirements of the action. Some actions, such as gross posture, re-

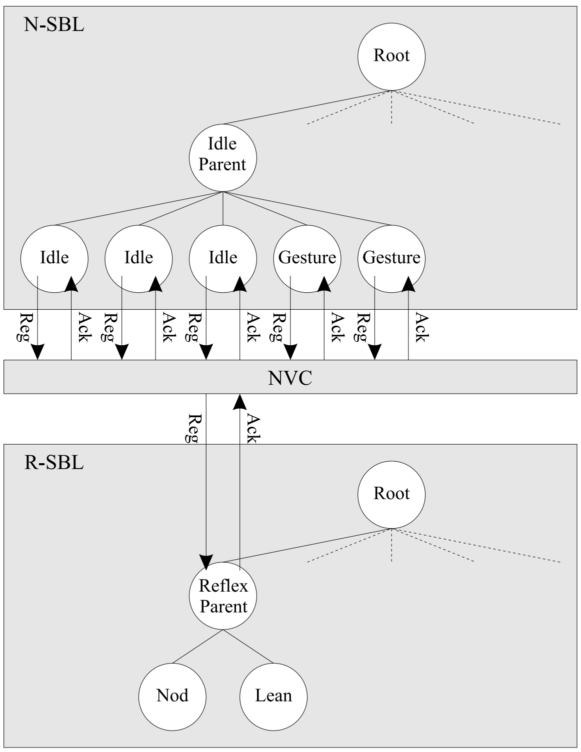

Figure 10: Special schemas register with the NVC object in a handshaking procedure. An acknowledgement from NVC is required because the SBLs do not know when the NVC object is ready to accept messages, so they attempt to register until successful. Within the R-SBL, the parent registers instead of the individual schemas, to preserve as closely as possible the mode of operation of the pre-existing conversational reflex system. The registration system also supports allowing schemas to deliberately change type at any time (e.g. from an idle schema to a behavioral schema) though no use of this extensibility has been made to date.

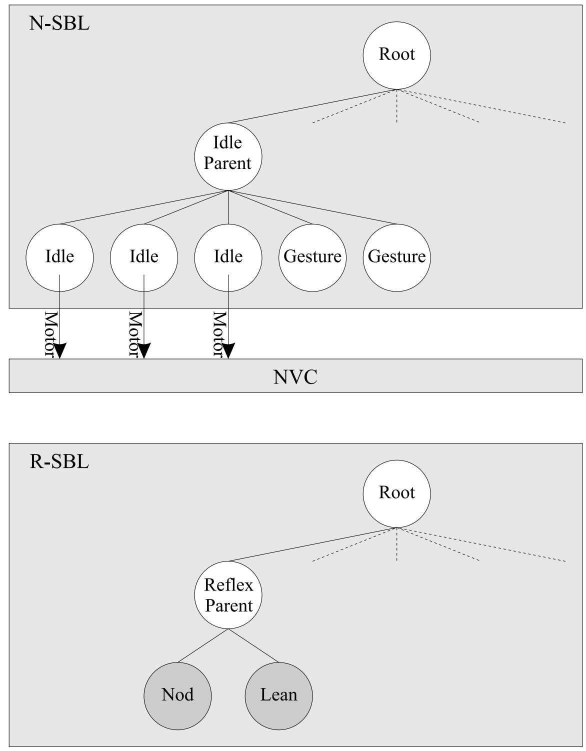

Figure 11: Normal idle activity of the system in the absence of gesture or reflex triggering. N-SBL idle and gesture schemas are active if resources permit. Active idle schemas send null motor commands to NVC on each time step. The dialog reflex parent is always active, but the individual dialog reflex schemas are inactive.

quire only partial control of a resource, whereas others require its total control. Partial resource management has just been described above, and it functions identically with commands that are also synchronous and originate in the R-SBL. However there are also non-verbal communication behaviors that require total resource control, such as emblematic gestures, and these are also implemented with the use of specialized "triggered" schemas: at the N-SBL level these schemas are called gesture schemas, and at the R-SBL level they are called dialog reflex schemas.

Gesture schemas reside within the same subtree of the N-SBL as the idle schemas; unlike the idle schemas, however, they do not attempt to remain active when no instrumental behavior is operating. Instead, they await gesture trigger messages from the NVC object, because the NVC object can not create motion commands directly, it can only modify them. Upon receiving such a message, a gesture schema de-

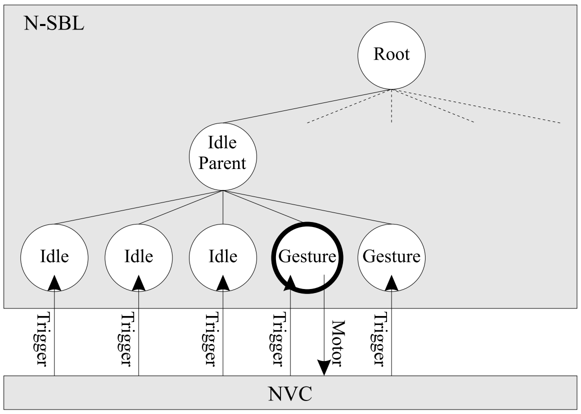

Figure 12: Triggering of a gesture occurs by NVC sending a trigger message to all registered N-SBL schemas. There is no distinction between idle and gesture schemas as idle schemas can include gesture-like responses such as postural shifts. Individual active schemas decode the trigger message and choose whether or not to respond with a motor command; in this typical example one gesture schema responds.

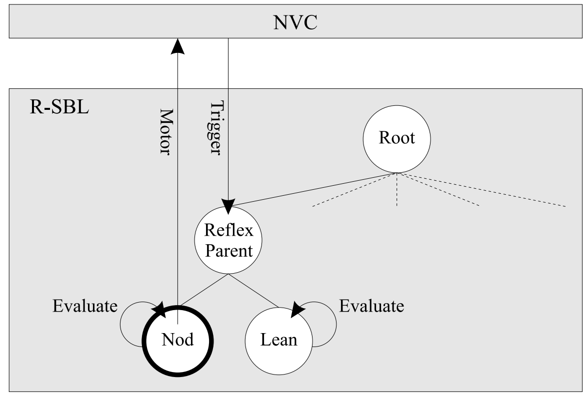

Figure 13: Triggering of a dialog reflex occurs by NVC sending an activation message to the active dialog reflex parent, which then signals the self-evaluation functions of its children. Children self-evaluating to active (e.g. speech is occurring) raise their own activation levels and send motor commands if required. The dialog reflex system remains active and periodically self-evaluating until receiving a deactivation message from NVC, so may either be tightly synchronized to non-causal data (e.g. pre-marked outgoing speech) or generally reactive to causal data (e.g. incoming speech).

termines if it is designed to execute the requested gesture; if so, it increases its activation level (AL), sends the motion command, and then reduces its AL again (Figure 12). The gesture, which may be an unparameterized motion from a predesigned set, is then executed by the MC object as usual.

Dialog reflex schemas operate in a similar but slightly different fashion. Existing research on QRIO has already resulted in a successful reactive speech interaction system that inserted contingent attentive head nods and speech filler actions into the flow of conversation(Aoyama, K. and Shimomura, H. 2005). It was of course desirable for the NVC system to complement, rather than compete with, this existing system. The prior system used the “intention” mechanism to modify the ALs of dialog reflex schemas, residing in the R-SBL, at appropriate points in the dialog. The NVC object manages these reflexes in almost exactly the same way.

Dialog reflex schemas are grouped under a parent schema which has no resource requirements and is always active. Upon receiving a control message from the NVC object, the parent schema activates its children’s monitor functions, which then choose to increase their own ALs, trigger the reflex gesture and then reduce their ALs again. Thus the only practical difference from the previous mechanism is that the reflex schemas themselves manage their own AL values, rather than the triggering object doing it for them. In addition to the existing head nodding dialog reflex, a second dialog reflex schema was added to be responsible for leaning the robot’s torso in towards the speaker (if the state of the robot’s interest and the Relationship deem it appropriate) in reaction to speech (Figure 13).

5.3 Execution Flow

The overall flow of execution of the NVC behavioral overlay system (Figure 14) is thus as follows:

- At system startup, the idle, and gesture schemas must register themselves with the NVC object to allow subresource allocation and communication. These schemas set their own AL high to allow them to send messages; when they receive an acknowledgement from NVC, they reduce their ALs to a level that will ensure that instrumental behaviors take priority (Figure 10).
- Also at system startup, the dialog reflex schema parent registers itself with NVC so that NVC knows where to send dialog reflex trigger messages (Figure 10).

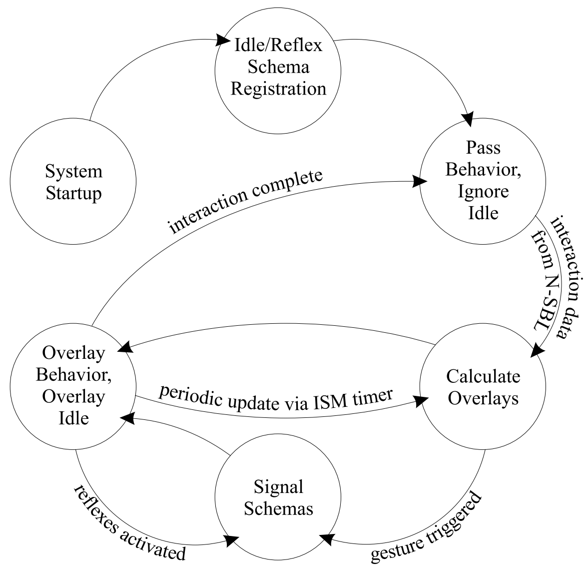

Figure 14: Execution flow of the NVC overlay system. Timing and synchronization issues such as responding to changes in the internal model and proxemic state (from STM) and waiting for the human to respond to an emblematic gesture are implicitly managed within the overlay calculation state (i.e. entering this state does not guarantee that any change to the overlays will in fact be made).

- Proxemic and body language activity commences when an N-SBL schema informs NVC that an appropriate interaction is taking place. Prior to this event NVC passes all instrumental motion commands to MC unaltered and discards all idle commands (Figure 11). The N-SBL interaction behavior also passes the ID of the human subject of the interaction to NVC so it can look up the appropriate Relationship.

- NVC generates overlay values using a set of functions that take into account the IM and Attitude of the robot, the Relationship with the human and the proxemic hint of the interaction (if available). This overlay generation phase includes the robot's selection of the ideal proxemic distance, which is converted into a walking overlay if needed.

- Idle commands begin to be processed according to the overlay values. As idle commands typically need only be executed once until interrupted, the NVC object maintains a two-tiered internal subresource manager in which idle commands may hold a resource until requested by an behavior command, but completing behavior commands free the resource immediately.

- To make the robot's appearance more lifelike and less rigid, the regular ISM update is used to trigger periodic re-evaluation of the upper body overlays. When this occurs, resources held by idle commands are freed and the next incoming idle commands are allowed to execute the new overlays.

- STM updates are monitored for changes in the proxemic state caused by movement of the target human, and if necessary the overlays are regenerated to reflect the updated situation. This provides an opportunity for triggering of the robot's emblematic gestures concerned with manipulating the human into changing the proxemic situation, rather than the robot adjusting it directly with its own locomotion. A probabilistic function is executed that depends mainly on the robot's IM and Attitude and the prior response to such gestures, and according to the results an emblematic gesture of the appropriate nature and urgency (e.g. beckoning the human closer, or waving him or her away) is triggered. Trigger messages are sent to all registered idle and gesture schemas, with it up to the schemas to choose whether or not to attempt to execute the proposed action (Figure 12).

- Activation messages are sent from NVC to the dialog reflex schemas when dictated by the presence of dialog markers or the desire for general reactivity of the robot (Figure 13).

6 Results

All of the non-verbal communication capabilities set out in Section 2 were implemented as part of the overlay system. To test the resulting expressive range of the robot, we equipped QRIO with a suite of sample Attitudes and Relationships. The Attitude selections comprised a timid, introverted QRIO, an aggressive, agonistic QRIO, and an "average" QRIO. The example Relationships were crafted to represent an "old friend" (dominant attribute: high Closeness), an "enemy" (dominant attribute: low Attraction), and the "Sony President" (dominant attribute: high Status). The test scenario consisted of a simple social interaction in which QRIO notices the presence of a human in the room and attempts to engage him or her in a conversation. The interaction schema was the same across conditions, but the active Attitude and Relationship, as well as QRIO's current EM and ISM state, were concurrently communicated non-verbally. This scenario was demonstrated to laboratory members, who were able to recognize the appropriate changes in QRIO's behavior. Due to the high degree of existing familiarity of laboratory members with QRIO, however, we did not attempt to collect quantitative data from these interactions.

Due to the constraints of our arrangement with Sony to collaborate on their QRIO architecture, formal user studies of the psychological effectiveness of QRIO's non-verbal expressions (as distinct from the design of the behavioral overlay model itself) have been recommended but not yet performed. In lieu of such experiments, the results of this work are illustrated with a comparison between general interactions with QRIO (such as the example above) in the presence and absence of the overlay system. Let us specify the baseline QRIO as equipped with its standard internal system of emotions and instincts, as well as behavioral schema trees for tracking of the human's head, for augmenting dialog with illustrators such as head nodding and torso leaning, and for the conversation itself. The behavioral overlay QRIO is equipped with these same attributes plus the overlay implementation described here.

When the baseline QRIO spots the human, it may begin the conversation activity so long as it has sufficient internal desire for interaction. If so, QRIO commences tracking the human's face and responding to the human's head movements with movements of its own head to maintain eye contact, using QRIO's built-in face tracking and fixation module. When QRIO's built-in auditory analysis system detects that the human is speaking, QRIO participates with its reflexive illustrators such as head nodding, so the human can find QRIO to be responsive. However, it is up to the human to select an appropriate interpersonal distance — if the human walks to the other side of the room, or thrusts his or her face up close to QRIO's, the conversation continues as before. Similarly, the human has no evidence of QRIO's desire for interaction other than the knowledge that it was sufficiently high to allow the conversational behavior to be initiated; and no information concerning QRIO's knowledge of their relationship other than what might come up in the conversation.

Contrast this scenario with that of the behavioral overlay QRIO. When this QRIO spots the human and the conversation activity commences, QRIO immediately begins to communicate information about its internal state and the relationship it shares with the human. QRIO begins to adopt characteristic body postures reflecting aspects such as its desire for interaction and its relative status compared with that of the human — for some examples see Figure 15. If the human is too close or too far away, QRIO may walk to a more appropriate distance (see Figure 3 for specifics concerning the appropriate interaction distances selected for QRIO). Alternatively, QRIO may beckon the human closer or motion him or her away, and then give the human a chance to respond before adjusting the distance itself if he or she does not. This may occur at any time during the interaction — approach too close, for example, and QRIO will back away or gesticulate its displeasure. Repeated cases of QRIO's gestures being ignored reduces its patience for relying on the human's response.

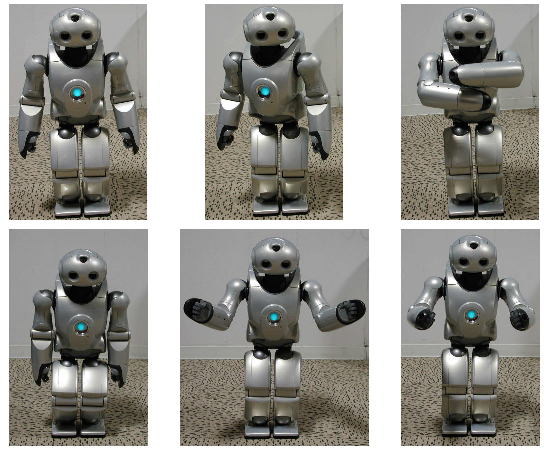

Figure 15: A selection of QRIO body postures generated by the overlay system. From left to right, top to bottom: normal (sociopetal) standing posture; maximum sociofugal axis; defensive arm crossing posture; submissive attentive standing posture; open, outgoing raised arm posture; defensive, pugnacious raised arm posture.

The behavioral overlay QRIO may also respond to the human's speech with participatory illustrators. However its behavior is again more communicative: if this QRIO desires interaction and has a positive relationship with the human, its nods and torso leans will be more pronounced; conversely, when disengaged or uninterested in the human, these illustrators will be attenuated or suppressed entirely. By adjusting the parameters of its head tracking behavior, the overlay system allows QRIO to observably alter its responsiveness to eye contact. The speed of its movements is also altered: an angry or fearful QRIO moves more rapidly, while a QRIO whose interest in the interaction is waning goes through the motions more sluggishly. All of this displayed internal information is modulated by QRIO's individual attitude, such as its extroversion or introversion. The human, already trained to interpret such signals, is able to continuously update his or her mental model of QRIO and the relationship between them.

It is clear that the latter interaction described above is richer and exhibits more variation. Of course, many of the apparent differences could equally well have been implemented through specifically designed support within the conversational interaction schema tree. However the overlay system allows these distinguishing features to be made available to existing behaviors such as the conversation schema, as well as future interactions, without such explicit design work on a repeated basis.

For the most part, the overlays generated within this implementation were able to make use of continuous-valued mapping functions from the various internal variables to the postural and proxemic outputs. For example, a continuous element such as the sociofugal/sociopetal axis is easily mapped to a derived measure of interaction desire (see Section 4). However, there were a number of instances in which it was necessary to select between discrete output states. This was due to the high-level motion control interface provided to us, which simplified the basic processes of generating motions on QRIO and added an additional layer of physical safety for the robot, such as prevention of falling down. On the other hand, this made some capabilities of QRIO's underlying motion architecture unavailable to us. Some expressive postures such as defensive arm crossing, for instance, did not lend themselves to continuous modulation as it could not be guaranteed via the abstraction of the high-level interface that the robot's limbs would not interfere with one another. In these cases a discrete decision function with a probabilistic component was typically used, based on the variables used to derive the continuous postural overlays and with its output allowed to replace those overlay values directly.

In addition to discrete arm postures, it was similarly necessary to create several explicit gestures with a motion editor in order to augment QRIO's repertoire, as the high level interface was not designed for parameterization of atomic gestures. (This was presumably a safety feature, as certain parameterizations might affect the robot's stability and cause it to fall over, even when the robot is physically capable of executing the gesture over the majority of the parameter space.) The gestures created were emblematic gestures in support of proxemic activity, such as a number of beckoning gestures using various combinations of one or both arms and the fingers. Similar probabilistic discrete decision functions were used to determine when these gestures should be selected to override continuous position overlays for QRIO to adjust the proxemic state itself.

QRIO's high-level walking interface exhibited a similar trade-off in allowing easy generation of basic walking motions but correspondingly hiding access to underlying motion parameters that could otherwise have been used to produce more expressive postures and complex locomotion behaviors. Body language actions that QRIO is otherwise capable of performing, such as standing with legs akimbo, or making finely-tuned proxemic adjustments by walking along specific curved paths, were thus not available to us at this time. We therefore implemented an algorithm that used the available locomotion features to generate proxemic adjustments that appeared as natural as possible, such as uninterruptible linear walking movements to appropriate positions. The display behaviors implemented in this project should thus be viewed as a representative sample of a much larger behavior space that could be overlaid on QRIO's motor output at lower levels of motion control abstraction.

7 Conclusions and Future Work

This research presented the model of a behavioral overlay system and demonstrated that an implementation of the model could successfully be used to modulate the expressiveness of behaviors designed without non-verbal communication in mind as well as those specifically created to non-verbally support verbal communication (e.g. dialog illustrators). Despite limitations in the motion controller's support for motion parameterization, a spectrum of display behaviors was exhibited. Future extensions to the motion controller could support increased flexibility in the communicative abilities of the system without major alterations to the basic overlay implementation.

The proxemic component of the behavioral overlay system emphasized that the spatial nature of non-verbal communication through the management of interpersonal distance is not only an important feature of human-robot interaction that is largely yet to be explored, but can also be effectively modulated by means of an overlay. The proxemics in this case were computed simply in terms of egocentric linear distances, so there remains plenty of potential for exploring alternative conceptualizations of proxemic overlays, such as force fields or deformable surfaces. In order to be truly proxemically communicative, future work is required in giving QRIO a better spatial understanding of its environment, not only including physical features such as walls and objects, but also a more detailed model of the spatial activities of the humans within it.

The overlay generating functions used in this project were effective display generators by virtue of being hand-tuned to reflect the general body of research into human postures, gestures and interpersonal spacing, with accommodations made for the differences in body size and motion capability of QRIO. As such, they represent a system that is generically applicable but inflexible to cultural variations and the like. The overlay generation components thus offer an opportunity for future efforts to incorporate learning and adaptation mechanisms into the robot's communicative repertoire. In particular, it would be beneficial to be able to extract more real-time feedback from the human, especially in the case of the proxemics. For example, given the nature of current human-robot interaction, being able to robustly detect a human's discomfort with the proxemics of the situation should be an important priority.

This work introduced significant additional infrastructure concerning QRIO's reactions to its internal states and emotions and to specific humans, in the form of the Attitude and Relationship structures. The Relationship support assists in providing an emotional and expressive component to QRIO's memory, and the Attitude support offers the beginning of an individualistic aspect to the robot. There is a great deal of potential for future work in this area, primarily in the management and maintenance of relationships once instantiated; and in "closing the loop" by allowing the human's activity, as viewed by the robot through the filter of the specific relationship, to have a feedback effect on the robot's emotional state.

Finally, returning to first principles, this paper sought to argue that since it is well supported that non-verbal communication is an important mode of human interaction, it will also be a useful component of interaction between humans and humanoid robots, and that behavioral overlays will be an adequate and scalable method of facilitating this in the long term. Robust HRI studies are now necessary to confirm or reject this conjecture, so that future design work can proceed in the most effective directions.

8 Acknowledgements

The work described in this document was performed at Sony Intelligence Dynamics Laboratories (SIDL) during a summer research internship program. The authors wish to gratefully acknowledge the generous support and assistance provided by Yukiko Hoshino, Kuniaki Noda, Kazumi Aoyama, Hideki Shimomura, the rest of the staff and management of SIDL, and the other summer internship students, in the realization of this research.

References

Althaus, P., Ishiguro, H., Kanda, T., Miyashita, T. and Christensen, H.I.: 2004, Navigation for human-robot interaction tasks, *Proc. International Conference on Robotics and Automation*, Vol. 2, pp. 1894–1900.

Aoyama, K. and Shimomura, H.: 2005, Real world speech interaction with a humanoid robot on a layered robot behavior control architecture, *Proc. International Conference on Robotics and Automation.*

Bailenson, J.N., Blascovich, J., Beall, A.C. and Loomis, J.M.: 2003, Interpersonal distance in immersive virtual environments, *Personality and Social Psychology Bulletin* **29**, 819–833.

Beskow, J. and McGlashan, S.: 1997, OLGA — a conversational agent with gestures, *Proc. IJCAI-97 Workshop on Animated Interface Agents: Making Them Intelligent.*

Bethel, C.L. and Murphy, R.R.: 2006, Affective expression in appearance-constrained robots, *Proceedings of the 2006 ACM Conference on Human-Robot Interaction (HRI 2006)*, Salt Lake City, Utah, pp. 327–328.

Blankenship, V., Hnat, S.M., Hess, T.G. and Brown, D.R.: 2004, Reciprocal interaction and similarity of personality attributes, *Journal of Social and Personal Relationships* **1**, 415–432.

Bowlby, J.: 1969, *Attachment & Loss Volume 1: Attachment*, Basic Books.

Breazeal, C.: 2000, *Sociable Machines: Expressive Social Exchange Between Humans and Robots*, PhD thesis, Massachusetts Institute of Technology.

Breazeal, C.: 2003, Towards sociable robots, *Robotics and Autonomous Systems* **42**(3–4), 167–175.

Breazeal, C., Brooks, A.G., Gray, J., Hoffman, G., Kidd, C., Lee, H., Lieberman, J., Lockerd, A. and Chilongo, D.: 2004, Tutelage and collaboration for humanoid robots, *International Journal of Humanoid Robots* **1**(2), 315–348.

Brooks, A.G., Berlin, M., Gray, J. and Breazeal, C.: 2005, Untethered robotic play for repetitive physical tasks, *Proc. ACM International Conference on Advances in Computer Entertainment*, Valencia, Spain.

Byrne, D. and Griffit, W.: 1969, Similarity and awareness of similarity of personality characteristic determinants of attraction, *Journal of Experimental Research in Personality* **3**, 179–186.

Cassell, J. and Vilhjalmsson, H.: 1999, Fully embodied conversational avatars: Making communicative behaviors autonomous, *Autonomous Agents and Multiagent Systems* **2**, 45–64.

Chi, D., Costa, M., Zhao, L. and Badler, N.: 2000, The EMOTE model for effort and shape, *Proc. 27th Annual Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH '00)*, pp. 173–182.

Christensen, H.I. and Pacchierotti, E.: 2005, Embodied social interaction for robots, *in* Dautenhahn, K. (ed.), *Proceedings of the 2005 Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB-05)*, Hertfordshire, England, pp. 40–45.

Davies, M. and Stone, T.: 1995, *Mental Simulation*, Blackwell Publishers, Oxford, UK.

Dittmann, A.: 1978, The role of body movement in communication, *Nonverbal Behavior and Communication*, Lawrence Erlbaum Associates, Hillsdale.

Ekman, P. and Davidson, R.J.: 1994, *The Nature of Emotion*, Oxford University Press.

Fong, T., Nourbakhsh, I. and Dautenhahn, K.: 2003, A survey of socially interactive robots, *Robotics and Autonomous Systems* **42**, 143–166.

Fridlund, A.: 1994, *Human Facial Expression: An Evolutionary View*, Academic Press, San Diego, CA.

Fujita, M., Kuroki, Y., Ishida, T. and Doi, T.: 2003, Autonomous behavior control architecture of entertainment humanoid robot SDR-4X, *Proc. International Conference on Intelligent Robots and Systems*, Las Vegas, NV, pp. 960–967.

Giese, M.A. and Poggio, T.: 2000, Morphable models for the analysis and synthesis of complex motion pattern, *International Journal of Computer Vision* **38**(1), 59–73.

Gordon, R.: 1986, Folk psychology as simulation, *Mind and Language* **1**, 158–171.

Guye-Vuilleme, A., Capin, T.K., Pandzic, I.S., Thalmann, N.M. and Thalmann, D.: 1998, Nonverbal communication interface for collaborative virtual environments, *Proc. Collaborative Virtual Environments (CVE 98)*, pp. 105–112.

Hall, E.T.: 1966, *The Hidden Dimension*, Doubleday, Garden City, NY.

Hara, F., Akazawa, H. and Kobayashi, H.: 2001, Realistic facial expressions by SMA driven face robot, *Proceedings of IEEE International Workshop on Robot and Human Interactive Communication*, pp. 504–510.

Heal, J.: 2003, *Understanding Other Minds from the Inside*, Cambridge University Press, Cambridge, UK, pp. 28–44.

Jordan, P.W.: 2000, *Designing Pleasurable Products: an Introduction to the New Human Factors*, Taylor & Francis Books Ltd., UK.

Kanda, T., Hirano, T., Eaton, D. and Ishiguro, H.: 2004, Interactive robots as social partners and peer tutors for children, *Human-Computer Interaction* **19**, 61–84.

Kiesler, D.J.: 1983, The 1982 interpersonal circle: A taxonomy for complementarity in human transactions, *Psychological Review* **90**, 185–214.

Knapp, M.: 1972, *Nonverbal Communication in Human Interaction*, Reinhart and Winston, Inc., New York.

Kopp, S. and Wachsmuth, I.: 2000, A knowledge-based approach for lifelike gesture animation, *Proc. ECAI-2000*.

Kopp, S. and Wachsmuth, I.: 2002, Model-based animation of coverbal gesture, *Proc. Computer Animation*.

Likhachev, M. and Arkin, R.C.: 2000, Robotic comfort zones, *Proceedings of SPIE: Sensor Fusion and Decentralized Control in Robotic Systems III Conference*, Vol. 4196, pp. 27–41.

Machotka, P. and Spiegel, J.: 1982, *The Articulate Body*, Irvington.

McCrae, R.R. and Costa, P.T.: 1996, Toward a new generation of personality theories: Theoretical contexts for the five-factor model, *in* Wiggins, J.S. (ed.), *Five-Factor Model of Personality*, Guilford, New York, pp. 51–87.

Nakauchi, Y. and Simmons, R.: 2000, A social robot that stands in line, *Proc. International Conference on Intelligent Robots and Systems*, Vol. 1, pp. 357–364.

Nass, C. and Lee, K.M.: 2001, Does computer-generated speech manifest personality? experimental test of recognition, similarity-attraction, and consistence-attraction., *Journal of Experimental Psychology, Applied* **7**(3), 171–181.

Orford, J.: 1986, The rules of interpersonal complementarity: Does hostility beget hostility and dominance, submission?, *Psychological Review* **93**, 365–377.

Pacchierotti, E., Christensen, H.I. and Jensfelt, P.: 2005, Human-robot embodied interaction in hallway settings: A pilot user study, *Proceedings of the 2005 IEEE International Workshop on Robots and Human Interactive Communication*, pp. 164–171.

Peters, R.A.II, Campbell, C.C., Bluethmann, W.J. and Huber, E.: 2003, Robonaut task learning through teleoperation, *Proc. International Conference on Robotics and Automation*, Taipei, Taiwan, pp. 2806–2811.

Reeves, B. and Nass, C.: 1996, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, Cambridge University Press, Cambridge, England.

Rose, C., Cohen, M. and Bodenheimer, B.: 1998, Verbs and adverbs: Multidimensional motion interpolation, *IEEE Computer Graphics and Applications* **18**(5), 32–40.

Sawada, T., Takagi, T., Hoshino, Y. and Fujita, M.: 2004, Learning behavior selection through interaction based on emotionally grounded symbol concept, *Proc. IEEE-RAS/RSJ Int'l Conf. on Humanoid Robots (Humanoids '04)*.

Smith, C.: 2005, *Behavior adaptation for a socially interactive robot*, Master's thesis, KTH Royal Institute of Technology, Stockholm, Sweden.

te Boekhorst, R., Walters, M., Koay, K.L., Dautenhahn, K. and Nehaniv, C.: 2005, A study of a single robot interacting with groups of children in a rotation game scenario, *Proc. 6th IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA 2005)*, Espoo, Finland.

Thorisson, K.R.: 1996, *Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills*, PhD thesis, MIT Media Laboratory, Cambridge, MA.

Walters, M.L., Dautenhahn, K., Koay, K.L., Kaouri, C., te Boekhorst, R., Nehaniv, C.L., Werry, I. and Lee, D.: 2005a, Close encounters: Spatial distances between people and a robot of mechanistic appearance, *Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots (Humanoids '05)*, Tsukuba, Japan, pp. 450–455.

Walters, M.L., Dautenhahn, K., te Boekhorst, R., Koay, K.L., Kaouri, C., Woods, S., Nehaniv, C.L., Lee, D. and Werry, I.: 2005b, The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment, *Proceedings of IEEE Ro-Man 2005, 14th Annual Workshop on Robot and Human Interactive Communication*, IEEE Press, Nashville, Tennessee, pp. 347–352.

Weitz, S.: 1974, *Nonverbal Communication: Readings With Commentary*, Oxford University Press.

Yamasaki, N. and Anzai, Y.: 1996, Active interface for human-robot interaction, *Proc. International Conference on Robotics and Automation*, pp. 3103–3109.

Yan, C., Peng, W., Lee, K.M. and Jin, S.: 2004, Can robots have personality? An empirical study of personality manifestation, social responses, and social presence in human-robot interaction, *54th Annual Conference of the International Communication Association*.

Zecca, M., Roccella, S., Carrozza, M.C., Cappiello, G., Cabibihan, J.-J., Dario, P., Takanobu, H., Matsumoto, M., Miwa, H., Itoh, K. and Takanishi, A.: 2004, On the development of the emotion expression humanoid robot WE-4RII with RCH-1, *Proc. IEEE-RAS/RSJ Int'l Conf. on Humanoid Robots (Humanoids '04)*, Los Angeles, California.