Humanoid Upper Torso Complexity for Displaying Gestures

Regular Paper

Robert Richardson, David Devereux, Jennifer Burt and Paul Nutter

School of Computer Science, University of Manchester, Manchester, UK
* Corresponding author E-mail: Rob@cs.man.ac.uk

Originally published in the Journal of Humanoids, ISSN 1996-7209

DOI: 10.5772/50919

© 2012 Richardson et al.; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract Body language is an important part of human-to-human communication; therefore body language in humanoid robots is very important for successful communication and social interaction with humans. The number of degrees of freedom (d.o.f) necessary to achieve realistic body language in robots has been investigated. Using animation, three robots were simulated performing body language gestures; the complex model was given 25 d.o.f, the simplified model 18 d.o.f and the basic model 10 d.o.f. A subjective survey was created online using these animations, to obtain people’s opinions on the realism of the gestures and to see if they could recognise the emotions portrayed. It was concluded that the basic system was the least realistic, complex system the most realistic, and the simplified system was only slightly less realistic than the human. Modular robotic joints were then fabricated so that the gestures could be implemented experimentally. The experimental results demonstrate that through simplification of the required degrees of freedom, the gestures can be experimentally reproduced.

Keywords humanoid robot, kinematics, body language, emotion.

1. Introduction

Body language is the physical movement of body parts due to muscle activation that would not be required for normal function. Emotions are portrayed through both the trajectory of these movements and their duration. The common perception of body language is that it is purely subconscious and reveals underlying and, perhaps, hidden emotional states. However, often body language is displayed consciously to accentuate verbal communication or deliberately display a strong emotional response.

These voluntary actions, called gestures, are perceived as produced, to “say something” [1], and are not the same as emotional reactions. However, gesture and emotional behaviours are closely related and the word gesture is often used to indicate a movement whether intentional or subconscious. Interpretation of gestures is dependant upon social factors and can vary between cultures therefore they need to be analysed in context to show their true meaning [2].

Body language, especially gestures, will be essential for natural communication between humanoid robotic systems and humans. Accidental or poorly exhibited body language may result in the breakdown of communication or even a feeling of dislike or fear by
humans. Gestures are suited towards being displayed by robotic systems as they are well bounded, with the action having a clear start and finish.

Automatic recognition of human body language from vision is a rapidly growing research area, but surprisingly little work has investigated the complexity of robot for displaying body language.

Several research institutions are developing humanoid robots. Asimo, is a 4ft tall, walking bipedal robot being developed by the American Honda Motor Co. [3]. It has 26 degrees of freedom, not including the 5 bending fingers on each hand. The SDR-4X from Sony, is a small humanoid built for entertainment. It has 38 DOF in total. The robot can recognise faces and speech. There has been consideration of the body language of the SDR-4X; extra degrees of freedom were added in the head and wrist to improve the expression of the robot [4]. However, it has no capability to bend at the wrist limiting the emotions that can be expressed. Waseda University in Tokyo are researching into several humanoid robots; Robita, Wabian, Wendy and Wamoeba-2R. Wamoeba-2R is the only robot capable of displaying body language in its arms, however it is not able to move its shoulders or trunk, therefore it is likely the ‘emotional’ experiences reported [5,6] are as a result of facial features or speech synthesis. The Massachusetts Institute of Technology have performed significant work in the area of socially interactive robots. Kismet is a robot formed with the image of just a head and neck. The main focus of the work on Kismet is to make it natural and expressive. The areas of research include facial expression, body posture and social cues. Head and eye orientation and facial expression are used as nonverbal signals to portray emotions [7,8]. MIT’s ‘Cog’, has a head, torso and arms but no legs. It has 22 degrees of freedom, similar to a human. Body language has been implemented on Cog, however it is again difficult to assess the realism of the body language when combined with facial expression. Table 1 summarises the humanoid robots discussed.

No conclusive work has demonstrated the importance of robotic body language for ease of human communication and the robots under development have different degrees of freedom and movement ranges. A robot designed specifically for natural human communication must have the capability to display emotion; however each additional joint of a robotic system adds significant cost. It is unlikely that the full complexities of human joints are required to be duplicated by a robotic system in order to display basic emotional responses.

This study seeks to gain a greater understanding of the required complexity of a robotic system in order to display emotion through animation. In section 2, three robot kinematic configurations are presented; the first with complexity approaching that of a human, the second with reduced complexity and the third with a very basic configuration. The implementation of these configurations as animations is discussed. Then 6 different gestures with well-defined and understood meanings are selected and animations are developed in section 3. Section 4 describes an internet survey performed to assess people’s ability to perceive the displayed emotions. Section 5 describes experimental development of the humanoid upper torso and section 6 illustrates single arm movement. Section 7 implements the gestures experimentally on the 10 d.o.f configuration and finally section 8 draws conclusions from the work.

| Robot    | d.o.f          | Purpose                                      |
|----------|----------------|----------------------------------------------|
| Asimo    | 26 + fingers   | Walking and balancing; function in the human living space |
| Robita   | 19             | Natural conversation system                  |
| Wabian   | 35             | Dynamic walking; collaborative work with humans |
| Wendy    | 52             | Advance dexterity and human interaction ability |
| Wamoeba-2R | 14             | Achieve emotion equal to that of humans to enhance the communication ability |
| H6       | 35             | Bipedal walking                              |
| Saika    | 12             | Skilful manipulations; behaviour-based movement control; modular |
| SDR-4X   | 38             | Entertainment                                |
| Biomorphic Arm | 9 + 3 per digit | Create a humanoid arm with the same degrees of freedom as a human arm; telepresence applications |
| Robonaut | 43             | Dextrous work in space                       |
| Valerie  | 111            | Domestic chores                              |
| Kismet   | 15             | Natural infant-caretaker interaction         |
| Cog      | 22             | Artificial Intelligence; humanoid interactions |

Table 1. Summary of current humanoid robots

2. Kinematic configurations

Three robot configurations were used in this study (figure 1).

Only the main joints have been analysed, for example the rib cage moves up and down in humans for some emotional states. No attempt has been made to represent these subtleties of motion. The study deliberately ignores facial or finger gestures as these are the dominant component of body language and their implementation would limit assessment of the limb movements. Each of the joints has a movement range similar to that of humans [9] irrespective of the system complexity.
The first robotic system (complex) approaches the complexity of a human with 25 degrees of freedom. Movements of the shoulder joint were limited to 5 degrees of freedom (d.o.f) and the vertebra in the back was limited to 3 d.o.f when in reality each vertebra has multiple degrees of freedom. The second system (simplified) removes some of the degrees of freedom such as the clavicle, reducing the degrees of freedom to 18 d.o.f. Finally, the third system (basic) reduces the d.o.f to 10.

Figure 1. Three robot configurations

Initially, the three robotic systems were developed in animation. Although the robots have different degrees of freedom, their outward appearance was identical. Figure 2 illustrates a static pose of the animation. The animations were constructed without facial features or muscular shapes to try and ensure they produce no strong emotion without motion. Grey was selected as a neutral colour scheme.

Figure 2. The outward appearance of the animations

3. Gestures

Six emotional responses were selected as they have a good range of movements allowing a wide range of emotions to be portrayed. Furthermore, the gestures are not alike in meaning or in movement. Therefore, the chances of confusing the gestures are reduced. The gestures are described in Table 2 [10],[11].

Each robot figure was animated to implement these movements within the restrictions of the d.o.f. Ideally, the eyes should lead the movement closely followed by the head, to suggest that it is the thoughts of the character that are driving its actions. In this situation the animation has no eyes therefore, it is very important that the head leads. How much the head lead by depends upon how much thought is going in to the action. When a character is happy, the body movements it makes are fast; the body movements of a sad character are slower and the head hangs down [12]. Subtle differences in motion can affect the believability of characters [13].

For each joint there are two types of human movement; ballistic movements and controlled movements. Ballistic movements are prepared in advance without any adjustments in motor control. Controlled movements are made at a moderate speed and are subject to change; they are amended during the movement, using feedback information [14]. The animations use ballistic movements as they are displaying emotional thoughts, hence they are more innate, already known, and will not be subject to changes.

People do not move symmetrically, therefore, to increase the realism, the right arm is made to move slightly before the left arm, which indicates right-handedness.

Trajectory paths were formed for each joint of the three configurations, for each emotion, over time. For example the complex system had 25 trajectory paths, one for each d.o.f. The trajectory was created by considering a few specific
points and interpolating the data points in-between. Figures 3-8 illustrate the movement of the animations made. For brevity only 3 frames of each movement are shown without reference to timing information. Typically each motion lasts for around 2 seconds. The animations show the complex (A), simplified (B) and basic (C) animation frames. It is apparent from examining the animation frames that the basic system is unable to realistically display some of the movements such as arm cross.

| Emotion            | Description                                                                                                                                 |
|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| 1) Akimbo          | The akimbo gesture is putting hands on hips. It is a confident and aggressive gesture, showing that the individual is prepared to “take steps”. It is an aggressive gesture because it is used to make the person look bigger. Also the palms of the hands are facing down, which shows dominance and confidence. The akimbo animation also includes a slight head tilt to the side as a questioning gesture – “what do you think you’re doing?” |
| 2) Arm cross       | This is a defensive gesture. It is guarded and shows disagreement, dislike, arrogance or anxiety. It is used as a barrier to block out undesirable circumstances. The spine twists to angle away from the person they are facing, showing negative feelings. The head is moved slightly so that it is still facing forwards. |
| 3) Hand behind head| The hand behind head gesture is negative. It is also known as the ‘pain-in-the-neck’ gesture. It is usually indicative of feelings such as uncertainty, frustration, anger or dislike. The gesture includes gazing down and angling the body away to represents feelings of defeat, guilt, or shame. |
| 4) Shrug display   | The shoulder shrug display is a global body movement, including not only the shoulders, but also the head, elbows, hands and torso. This is a submissive gesture, showing uncertainty, resignation or helplessness. The movements are: the shoulders are raised, head tilted sideways, elbows bent and held in, palms shown and upraised, body bent forwards at the waist. |
| 5) Dominance       | Having the palm of the hand facing downwards shows dominance, confidence, assertiveness and authority. Moving the hand up and down (beating) is symbolically beating the listener into submission. The head tilting backwards shows superiority, arrogance and disdain. Having the other hand on the hip helps to confirm the attitude of confidence, it too makes the palm face down |
| 6) Excited         | The gestures included in the Excited movement are nodding the head and rubbing the palms of the hands together. Nodding the head up and down is affirmative, showing understanding, approval or agreement; emphatic nods show feelings of conviction, excitement or sometimes rage. This excited movement has emphatic nods. Rubbing the hands together shows positive expectation; quickly rubbing together shows excitement. |

Table 2. Gesture descriptions

4. Survey to assess animations emotional expression

A survey was performed to gain some incite into the realism of each gesture and to compare the different robot configurations.

The survey contained all the animations displaying ‘emotional’ states and each reviewer was asked the following questions:

“What do you think the figure is feeling or thinking?”
“What do you feel or think when you look at the figure?”

Figure 3. Akimbo animations

Figure 4. Arm cross animations

Figure 5. Hand behind head
The people performing the survey were unaware that the animations had different d.o.f and the animations were presented in a random order, so that the animations with the same gesture could not be directly compared. The complex animation has the greatest d.o.f therefore, it is likely that it would receive the highest score, conversely the simplified animation would have a lower score and the basic would have the lowest score. Nineteen anonymous responses received through the Internet survey were examined. To be labelled as correct, the emotion given had to describe the general emotional area, since emotions are very subjective and the movements that people make when experiencing them are very individual.

The results of the survey for the recognized emotion are shown in Table 3. The body language of the basic animation is the least recognisable, with only 34% of the movements being identified. The complex animation has the most correct responses and therefore has the most recognisable movements.

Table 3. Survey results

| Gesture               | Basic (%) | Simp. (%) | Comp. (%) | Overall |
|-----------------------|-----------|-----------|-----------|---------|
| Hand Behind Head      | 8         | 31        | 36        | 25      |
| Excited               | 47        | 67        | 47        | 54      |
| Arm Cross             | 13        | 56        | 89        | 53      |
| Dominance             | 50        | 63        | 63        | 58      |
| Akimbo                | 31        | 87        | 89        | 69      |
| Shoulder Shrug        | 57        | 94        | 94        | 82      |
| Average               | 34        | 66        | 70        | 57      |

Overall, the least recognisable gesture was Hand Behind Head; the most recognisable gesture overall was the Shoulder Shrug. Although the complex animation represented the gestures with the greatest accuracy, the simplified robot was almost as recognizable. Some of the responses to the survey indicate an emotional response is being perceived:

- It is commending me for something
- I feel he’s angry with me
- We are beginning to work as a team
- reminds me of my wife in a bad mood

These results indicate that the clavicle joint in particular plays little part in gesture representation. Reducing the d.o.f to 10, vastly reduces the clarity of emotional states displayed. Little movement of the wrist was implemented on any of the configurations and the complexity of the neck on the simplified configuration was not required for the majority of the movements. Therefore, the structure of the upper torso was reduced to that shown in figure 9 for experimental implementation.
5. Construction of the upper limb gesture system

Constructing an arm from four degrees of freedom it is still a relatively complex task. Modularity is the best approach to keep the design simple and relatively affordable. Two different modular units were used to create the arm; with different torque and weight performance. Each unit consists of a single motor and potentiometer to allow precise joint angle control. The modules were designed to allow joint constructions for any serial/parallel combination of joints. This allows extremely versatile construction, with the drawback that spherical joints are modelled by three separate joints with offset; this results in only an approximate spherical joint. Table 4 describes the performance of the two modules and figure 10 illustrates the modular joint system in a pitch rotation configuration (around the x axis). The modules can also be connected in relative roll (around y) and yaw (around z) configurations. Each module can be connected directly end to end or spaced using hollow carbon fibre rod; this allows complex joints and structures to be implemented. The final kinematics of the constructed arm is shown in figure 11 and a photo of the system is shown in figure 12. The main issue with using modular single degree of freedom joints is apparent in the shoulder joint. Ideally this should be a single spherical joint with all the axes aligned. However, here only 2 of the joints are aligned resulting in translations that vary with joint orientation.

Figure 9. Degrees of freedom for experimental implementation

Table 4. Module joint specifications

| Joint Type | Description |
|------------|-------------|
| Large joint (joint 2) | Powered by Maxon A-max 32 with maximum output torque of 4.5Nm (limited by gearbox rating) |
| Small joint (all other joints) | Powered by Maxon A-max 22 with maximum output torque of 1Nm (limited by gearbox rating) |

Figure 10. Three modules connected in a relative pitch configuration

6. Gesture implementation

Comparative experiments were performed on a single arm to develop control algorithms; when two arms are used one arm should lead the other and the movements should be not entirely symmetrical to avoid looking ‘mechanical’. Therefore, it is easier to interpret the control performance of a single arm. The gestures were ‘taught’ to the arm by manually moving the arm in passive motion through a time varying series of motions. The arm then ‘replayed’ the motions in active mode through joint space PID controllers. The joints are defined (figure 11) as: joint 1 – rotation into and out of the plane around frame 1, joint 2 – Large joint that raises the arm against gravity, joint 3 – Spins the elbow, joint 4 – the elbow joint.
Although the animations contain two arms, head and torso movement, they offer a useful comparison against the performance of the simplified arm developed here. Figure 13 illustrates the akimbo single arm experimental response and repeated animation frames for ease of comparison. The illustrations show ‘key points’ of motion and do not necessarily correspond to sequential time slices. Figure 16 shows the full angle movements against time in joint space. The experimental arm is capable of accurately reproducing the akimbo action. Most joints are involved in the motion apart from the first joint. It is important to note that the motion needs to be considered in joint space as the full structural configuration expresses gestures, rather than the traditional robotic focus on end effector movement.

The dominance gesture is shown in figures 14 & 17. To perform the gesture the whole arm is rotated forward to allow it to be raised in line with the viewer. The rotation of the arm around joint 1 results in movement of the arm out of the page, due the shoulder joints not being coincident. This also aligns the large joint length ways, which looks ungainly. However, in general the gesture can be represented with reasonable accuracy. Note that to perform the gesture both arms perform different actions. The other arm is producing an Akimbo action, which has already been demonstrated to be producible on the system.

The shoulder shrug gesture is synonymous with raising the clavicle; this is a movement the arm is not capable of performing. Therefore, the gesture is expressed solely in the rest of the arm. Figures 15 & 18 illustrate the gesture being implemented. The arm creates the correct profile, however the gesture is not easily recognisable without the distinctive shoulder raise. It maybe that is specific contexts this gesture will be sufficient to be understandable.
7. Two arm experimental system

Following the successful implementation of the single arm, a full experimental system was constructed with the complexity shown in figure 9. The arm trajectories were defined as in the previous section however, one arm leads the motion and each arm motion was slightly different to avoid a ‘mechanical’ look to the motion.

Figure 19 shows motion frames of the akimbo gesture. Note that the right arm leads the left indicating a ‘right handed’ motion. The trajectory paths of the motions are also slightly different. Figure 20 illustrates the dominance gesture. This gesture is formed from different left and right arm motions. The left arm reaches towards the observer resulting in stronger response than in animation. Figure 21 illustrates the shrug motion, with the right arm again leading. Supple movements of the head add to the gesture effectiveness.

These results show that you do not need high degrees of freedom to display recognisable gestures.

8. Conclusions

This work has investigated the expression of emotion by the upperbody motion of humanoid robots. It has been demonstrated in both animation and experimentally that gestures can be displayed from robot arms with far less...
degree of freedom than humans. The reduced complexity has enabled an experimental system to be constructed with relative ease to implement these gestures. Further work will perform detailed interaction studies to determine the emotional response to the gestures and compare/contrast the differences between the emotion generated by animation and those from the experimental system.

9. References

[1] Kendon, A., 1996. An agenda for gesture studies. *Semiotic review of books*, 7 (3), 8-12.
[2] Birdwhistell, R.L., 1971. Kinesics and context: essays on body-motion communication. London, Allen Lane The Penguin Press.
[3] American Honda Motor Co., Inc., 2003. The Honda humanoid robot technical information [online]. American Honda Motor Co., Inc., Corporate Affairs and Communications, Torrance, California, January 2003. Available from: http://asimo.honda.com
[4] SONY, 2002. Sony develops small biped entertainment robot [online]. (Press Release March 19 2002). Sony Corporation, Tokyo, Japan. Available from: http://www.sony.net/SonyInfo/News/Press/200203/02-0319E/
[5] Ogata, T., Matsuyama, Y., Komiya, T., Ida, M., Noda, K., and Sugano, S., 2000. Development of emotional communication robot: WAMOeba-2R – experimental evaluation of the emotional communication between robots and humans. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems 2000* (IROS 2000). Proceedings, Vol. 1, 2000, 175-180.
[6] Sugano, S., 2002. Intelligent machine laboratory [online]. Sugano Laboratory, Department of Mechanical Engineering, Waseda University, Tokyo, Japan. Available from:

[7] Breazeal, C., 2000. Sociable machines: expressive social exchange between humans and robots. *Thesis (PhD)*, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, Massachusetts.
[8] Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, B., and Williamson, M., 1998. The Cog project: building a humanoid robot. In: C. Nehaniv, ed. *Computation for metaphors, analogy and agents*, Springer lecture notes in artificial intelligence 1998, Springer-Verlag, Vol. 1562.
[9] DEPARTMENT OF TRADE AND INDUSTRY, 1998. Adultdata: the handbook of adult anthropometric and strength measurements – data for design safety. UK, Department of Trade and Industry, URN 98/736.
[10] Givens, D.B., 2002. The nonverbal dictionary of gestures, signs & body language cues [online]. Center for Nonverbal Studies, Spokane, Washington. Available from: http://members.aol.com/nonverbal2/diction1.htm
[11] Pease, A., 1997. Body language: how to read others’ thoughts by their gestures. 3rd ed., reprinted 2002. London, Sheldon Press.
[12] Lasseter, J., 1994. Tricks to animating characters with a computer. In: *Animation tricks*, Course 1, SIGGRAPH 94.
[13] Hodgins, J. K., 1998. Animation lab [online]. The Graphics, Visualization and Usability Center, Georgia Institute of Technology. Available from: http://www.cc.gatech.edu/gvu/animation/index.html
[14] Kopp, S., and Wachsmuth, L., 2000. Planning and motion control in lifelike gesture: a refined approach. In: *Post-proceedings of computer animation 2000*. IEEE Computer Society Press, 2000, pp92-97.