Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Conversational Gestures in Human-Robot Interaction

Paul Bremner, Anthony Pipe, Chris Melhuish
Bristol Robotics Laboratory
Bristol, UK.

Mike Fraser, Sriram Subramanian
Computer Science Dept.
University of Bristol
Bristol, UK.

***Abstract*— The human sciences have demonstrated that gesture is a critical element of human communication. While existing graphical solutions are appropriate for virtual agents, solving arm trajectories for physically embodied robots requires that we consider the challenges of robot dynamics within a real-time gesture framework. We explore and evaluate a low computational-cost gesture production algorithm that can generate adequate gesture trajectories in a humanoid torso, as judged by participants in Human-Robot gesturing studies presented in this paper. Our approach produces a constrained inverse-kinematic solution for the start and end points, and generates appropriate wrist angles. Gesture time is used to calculate the joint accelerations to give a smooth, direct hand movement. Selecting open hand gestures as an example gesture sub-domain, we implement our controller on BERTI, a bespoke upper-torso humanoid robot (Fig. 2). A qualitative pilot study highlights gesture features salient to users: gesture shape, timing, naturalness and smoothness. A controlled experimental study then demonstrates that, by these metrics, our algorithm performs well; despite some dissimilarities with users' own gestures. We establish some salient points of robot gestures based on these studies.**

***Keywords*—Human-Robot Interaction, Conversational Gesture, Low-Cost Motion**

I. INTRODUCTION

Humanoid service robotics is a rapidly growing area of research. An important justification for the future presence of humanoid service robots is that they are able to carry out the same tasks as people, and in a similar manner. Ideally, it will be possible for them to be cooperated with, understood and trusted by a layperson, without having to adjust their behaviour or expectations. Research over the past decade (cf. [1]) has shown that anthropomorphic robots lead users to expect human-like behaviour. Thus, in order for a humanoid agent to engender the familiarity and trust desired, it must communicate in a human-like manner. A key facet of human communication is gesture and successful mimicry of this behaviour, at a suitable level of accuracy that we investigate in this paper, should help us to achieve the aims stated above [2].

There are numerous kinds of interactional uses of the body in communication, which may be broadly interpreted as 'gesture', but here we focus on movements in the upper torso, and particularly the arms and wrists, which accompany our everyday speech. It is commonly established that gestures, unlike the speech they accompany, have few concrete rules for their application or their form, beyond their sequential relevance to people within a particular interaction [3][4]. Furthermore, there is no global hierarchy or syntax for the combining of multiple gestures [4]. Evidence suggests that gestures are contingently designed rather than recalled, and that this is achieved in conjunction with producing speech to create a complete, relevant and coherent utterance [3]. However, although highly individual in practice, gesture production does have some structure, and it is likely that it is loosely produced from a shared schema [4]. Prior to attempting to understand sufficient detail of this schema to mimic it (such as is done by the BEAT system [5]) some key details need to be understood. Broadly speaking, these concern movement generation and user perceptions of a humanoid robot gesturing system.

The main contribution of our work is to provide some of the understanding needed to create a credible robot gesturing system. We have used a bespoke upper-torso humanoid robot to conduct this work. We have created a control scheme that can produce smooth arm motions using a computationally simple algorithm. We have then used this system to conduct two user studies into user expectations and perceptions of a robot gesturing system. As a result, we have identified recommendations for designing robot gesturing systems.

II. RELATED WORK

Both virtual graphical characters and embodied robotic characters may embody anthropomorphic properties such as gestures. There are, however, significant differences in gestural requirements between these approaches. Much work has been carried out concerning conversational gesture generation using embodied conversational agents (ECAs). A common approach for ECA motion generation is to use libraries of movements either designed by an expert animator [5], motion captured [6], or reproduced from finely analysed human movements [7]. These gesture primitives can be parameterised and combined to produce the required movements. However the range of possible gestures is limited by the size of the library and there is some difficulty in temporally adjusting them (to sync. with speech) while maintaining human-like motion [5].

To address these problems for ECAs, the MAX system of Kopp and Wachsmuth [8] generates realistic looking gesture movements from symbolic descriptions. A human-like hand trajectory is generated for the desired gesture and it is accurately followed using a kinematics approach. While these precise trajectories might be followed in a robotic system, it is far more computationally complex to do so as the dynamics of the embodied system must be solved in addition to the inverse kinematics. This makes achieving an embodied robot character

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

that produces gestures reactively in real-time a somewhat different and, in some ways, more difficult problem.

Previously, humanoid robots that use gesture with speech have focused on pointing gestures (deictics). Robovie (Ishiguro et al. [9]) and Melvin (Sidner et al. [10]) point to objects they are referring to in their speech to ensure attention sharing with a human user. In the work of Breazeal et al. [11], more complex mime gestures as well as deictics are produced; however, in that work gestures are the robot's sole means of communication. More recently there has been some work carried out on using some conversational gestures with humanoid robots. In the work of Bennewitz et al. [12] their robot (Fritz) uses some speech accompanying gestures along with a wider range of interactive behaviours. Their system appears well perceived by users, but how the different behaviours contribute to this is not investigated. Kim et al. [13] look at the plausibility of autonomously adding gestures to speech for a humanoid robot. However, they have yet to implement their system on an embodied robot and provide no user study to identify perceptions of robot gesturing, or requirements for user familiarity.

III. CONTROL ALGORITHM

There is considerable difficulty in producing a control scheme that accurately follows human-like arm (endpoint and joint) trajectory and velocity profiles for a robot. The complex set of movement-dynamics considerations involved, many of which are of a non-linear nature, require highly computationally intense solutions. In the future we want to be able to create gesture motions in real-time 'on the fly'. Therefore, we are motivated to find an alternative approach that has a low computational load, while still able to produce effective gestures. We contend that useful gestures can be produced under such a constraint, by mimicking key *features* of human gestural motion, as opposed to production of precise human-like motions. We suggest that the key features we need to imitate for effective gestures are correct end points and smooth, direct movement. We suggest smooth movement because theories of human arm trajectory formation (e.g. minimum jerk [14]) indicate that it is desired. We suggest end points because the direction of movement (not trajectory) is often used in gesture descriptions in the literature [3][4], further, we propose that when combined with a smooth, direct trajectory the desired direction will be perceived. This idea is further suggested from our empirical observations of human gestural movement. Consequently, we have produced a simple control scheme that achieves this; we have tested whether effective gestures are produced by this scheme in the HRI experiments detailed in the following sections.

A gesture is made up of three phases [3]: preparation, stroke and retraction. Appropriate movement time and end-point locations are required for each phase in order to calculate the required motor commands. Stroke time must match the word to be emphasised [3], while timing for the other phases is estimated based on sentence structure. The preparation phase is moving from rest point, *R*, to stroke start-point *S*. The stroke phase is moving from *S* to stroke end-point *P*. The retraction phase is moving from *P* back to *R* (see Fig. 1). The movement for each phase is triggered by hand coded events in the speech.

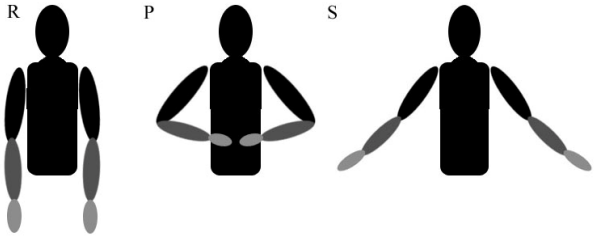

Figure 1. Examples of gesture end-points

To begin with, the inverse kinematics are solved for *R, P* and *S*. In order to constrain the calculations, the x components (horizontal when the robot is viewed from the front) of the wrist direction vectors - X_R, X_P and X_S, are used alongside the desired spatial location of the wrist. We have selected this method of constraint as the x components can be intuitively determined to generate the arm positions required. This calculation gives two possible solutions, but only the 'elbow down' solution is appropriate for a humanoid robot in this application. Next, wrist joint angles at P are calculated using a simple heuristic that requires approximate hand orientation relative to the elbow joint axis at P to be specified. Wrist rotation angle for the end-point is calculated to orientate the closer of the two wrist-joint axes perpendicular to the gesture direction, starting with the wrist rotated at the approximate rotation specified. The aligned joint is given a random value that moves the hand in the same direction as the gesture, while the other is set to zero. We do this to give the appearance of momentum of the hands which would require some effort to oppose. This idea is again based on our empirical observations of human movement; motion captured data of a person carrying out typical gestural movements often showed the suggested wrist deflections. In addition, the movements of the robot appeared 'stiffer' and less natural prior to its inclusion. The wrist joint angles at S are the same as those used when the robot is at R. Using a triangular velocity profile for each joint, accelerations are calculated so that all the joints will start and finish moving at the same time for each phase. This produces the desired smooth movement.

IV. STUDIES

A. Aims and Rationale

Our goal was to establish the extent to which our approach produces sufficiently credible gestures. We carried out a pilot study to identify the features of humanoid robot gesturing most salient to users. We then conducted a quantitative experiment to investigate whether the robot's gestures are successful according to these points, using their own gestures as a frame of reference.

B. Pilot Study

1) Set-up and Materials

The BERTI torso (Fig. 2) was designed and built by Elumotion Ltd., in conjunction with our laboratory. Each arm has seven degrees of freedom, and there are nine degrees of freedom in each hand, as well as two each at the waist and neck, giving a total of thirty-six. Each joint is actuated with a DC motor, coupled to a harmonic drive, and has a local motor controller commanded from the base PC using a CAN bus. The

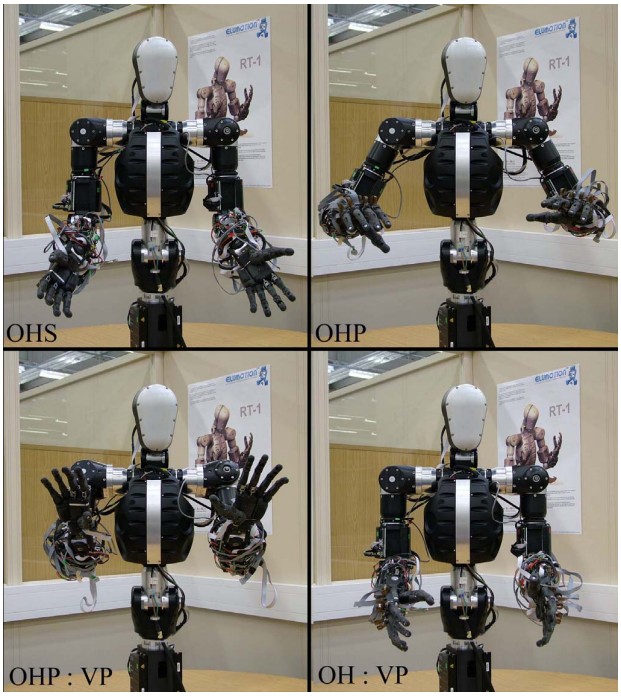

Figure 2. End points of open hand gestures performed by BERTI

arm joints are capable of speeds similar to that of human movement. However, due to mechanical limitations, the finger and wrist joints are not able to replicate human speeds; although they are capable of adequate ranges of motion for gesturing. Thus, BERTI is capable of moving in a human-like manner as well as having a human-like torso structure.

There are several commonly identified types of gesture used in different circumstances: deictic (pointing) gestures, metaphoric (that visually represent abstract concepts) and iconic gestures (that visually represent objects or motion), and beat gestures (used for emphasis and timing, rhythmically linked to accompanying speech) [4]. In order to ensure that the gestures produced were contextually correct, we chose to use a small range of well-understood gestures; specifically, gestures of the open hand as defined by Kendon [3]. If the gestures used were not correct, it would be likely to cloud the opinions of the gesture motion quality.

For each of the four types of open hand gesture to be used, a script of 3 different appropriate sentences was written. The stroke end point S for the four types of open hand gestures implemented are shown in Fig. 2:

- Open Hand Supine, lateral palm (OHS) – The hand is palm up moving from the centre of the torso outwards. Examples of use are: where something is declared obvious, presenting an idea, asking a question. e.g. 'I don't know, what do you think?"
- Open Hand Prone, lateral palm (OHP) – The hand is palm down and moves from the centre of the torso outwards. Examples of use are: when a line of action is cut off/suspended, no further action or exceptions possible, extreme assessment made. e.g. "The death-star was totally destroyed."
- Open Hand Prone: Vertical Palm (OHP:VP) – The hand is palm outwards and moves away from the body. Examples of use are: the speaker wishes a line of action to be halted, the speaker has halted a line of action. e.g. "He started but I stopped him."
- Open Hand: Vertical Palm (OH:VP) – The palm of the hand is facing across the front of the torso and moved down in a chopping motion. Examples of use are: specifying topics to be discussed, defining boundaries. e.g. "The time available is limited."

The input parameters are hand-coded, based on the requirements of the gesture to be generated. The few parameters that need to be specified can be done so intuitively making gesture scripting straightforward - an advantage of the simplicity of our control algorithm.

The BERTI robot performed the movements and the script was read using the Microsoft text to speech (TTS) engine. A TTS engine is used so that the speech is controlled across the experiment, thus eliciting a fair comparison. The Microsoft TTS engine is very simple and there are no inflections in the speech that might create unintended shifts in emphasis away from the gestures. Gesture timing with respect to the audio sentence playback was estimated utilising theory from the gesture analysis literature [3][4]. The gesture is deployed with the word that carries the related semantic meaning. The timing was further verified by observing a volunteer reading the sentences out loud while performing the type of gesture requested.

2) Method

7 participants, 4 male, 3 female, aged 26-58, took part in the pilot study. Each participant observed BERTI's performance of one of the example sentences for each of the four types of open hand gesture. Each sentence was repeated twice so that participants could form a confident opinion. Each sentence was chosen at random from the three available for each gesture type. Before the experiment began, participants were shown the robot and asked for their expectations of how the robot would behave when gesturing with speech. An open-ended questionnaire and informal interview was used to gather participants' opinions on the gestures for each trial.

3) Pilot Study: Results

All the participants expected the gestures to appear mostly correct but not perfect – errors in timing, jerky movement and so on were expected. This correlates well with our expectation that the robot's movements need not be perfectly human-like.

The questionnaires were analysed for trends in the opinions of the participants by extracting opinions shared by a majority, and these were summarised. The participants thought that, in the majority of trials, movement of the robot was smooth, with correct timing and gestures that seemed natural.

Detailed user opinions of the control system indicated that the gestures created were credibly 'human-like'. However, we wished to explore this idea in more detail, so we asked for

properties that our pilot study participants felt were salient to this judgment. The majority of responses concerning features of credible gesturing were *smoothness* of movement, *shape* of gesture, *timing*, and *naturalness* of the gesture's end point. These formed the basis for a more focused study.

C. Quantitative Experiment

1) Method

The primary goal of this study was to quantify whether the robot performs well in the metrics identified in the pilot study as being 'human-like', rather than a qualitative judgement of how well the robot performed. We did not want to simply quantify how much worse the robot was at gesturing than a human. Rather, we wanted to compare participants' ratings of the properties of the robot's gesture with their ratings of its similarity to their own gestures. We created this method to see whether the participants' opinions of the gestures were affected by the similarity to their own movements. Our goal was to create a ground-truth for gesture that was independent of a generalised anthropomorphic ideal and instead represented a gestural ideal for the robot. This should help us to establish whether the robot was performing better when judged as a gesturing robot than when compared to a human.

We used the same script and gestures as in the pilot. We chose two of the three sentences for each open hand gesture type for a total of eight sentences. 10 participants (5 female) aged 21-40 were recruited. None of them had taken part in the pilot study.

The first stage of the experiment was to record the participant reading the selected sentences on video. They were told that all the gestures were open palm and approximately what direction the palms of the hands should be facing (up, towards the camera, etc.) for each sentence. The shape and the timing of the gesture were left entirely up to the participant. Owing to the minimal direction given the gestures produced appeared natural.

Once the recording was complete, the procedure for the experiment was explained to the participant. The four metrics that they would be rating the robot on were described to ensure the correct qualities were assessed. The participant then watched the video of their performance for each sentence followed by the robot performing the same sentence twice. By watching themselves the participant got a frame of reference against which they could rate the robot (in addition to its use for the similarity comparison). They then rated the robot's performance on the metrics of smoothness, shape, timing and naturalness using a 5-point Likert scale (from -2 to +2). They were finally asked to rate the similarity between their own and the robot's gesture, again on a 5-point scale. The order of sentences was counterbalanced (partial Latin square) across the participants.

2) Results

The mean scores for the four gesture properties are shown in fig. 3 below. All participants rated the robot's gesture properties as above average, with an overall mean of 0.9 on a -2 to +2 rating scale. These ranged from 1.15 for gesture shape to 0.73 for smoothness. If we take 2 to represent a perfect gesture property, these results are promising. The participants'

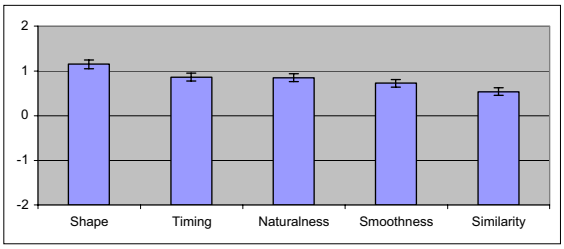

Figure 3. Ratings of Gesture Properties

judgement as to the overall gesture similarity to their own gesture is used to investigate whether there is a connection between it, and their gesture ratings. Participants rated the overall robot gesture at a mean of 0.54 against their own gesture (similarity in fig. 3), much lower than their assessments of the individual properties of the robotic gesture itself. A univariate analysis of variance (within subjects) over the data with users as random factors found a main effect of Gesture Property $F(4,36) = 3.686$, $p<0.05$.

Pairwise (Tamhane, unequal variance assumed) post-hoc comparisons found a significant difference between (Similarity, Shape) and (Smoothness, Shape). In addition, we conducted an ANOVA exploring differences between the ratings of the four open hand gesture types, and did not find any significant difference between the four open hand gesture types. $F(3, 27) = 2.045$, $p= 0.131$.

V. Discussion

The most important effect we found was a disparity between participants' comparisons of the robot's similarity with their own gesture, and their ratings of the properties of the robot's gesture itself. These numerical differences were reinforced by a significant statistical difference between their ratings of human-robot gestural similarity and robot gestural shape. We suggest that our innovative method of measurement has shown that high similarity (to human gesture) is not required for highly rated robotic gestures. This finding underpins and supports our novel approach of introducing a low computational-cost gesture production algorithm.

We have also contributed four properties of gestural production in robots and assessed their relative merits under our approach. It is clear that the shape of the gesture trajectory was rated as the most successful aspect of the algorithm. This is important, as it validates our hypothesis that precise replication of human hand trajectories is unnecessary in this application and validates the efficacy of our simple heuristics. The smooth stroke movement and generated wrist motion has created the desired impression on the users. While our findings hold true for the different gestures tested, further work is necessary to determine if they hold true for a wider range of gestures. However, the smoothness of the gestures was rated as significantly worse than the shape, and we will need to focus attention on this aspect of the trajectory production. Errors in the timing of transition between gesture phases are the likely cause of this issue. It was difficult to glean from the literature correct timing for the preparation and retraction phases (to and from rest) for a given gesture. Consequently, timings were guessed at the gesture coding stage and assumed to be

relatively unimportant. We now feel this is an issue that needs to be addressed for further work on robotic gesture generation. Although we have identified one area as worse than another, all of the gestural traits were rated as positive overall.

VI. CONCLUSIONS & FUTURE WORK

We have verified an algorithm that produces human-like movements in a low-cost manner for robotic open hand gesture production. Our findings indicate that the approach is feasible and useful, and that the gestures produced are more highly rated than direct comparison with human gesture might suggest.

A video of examples of each of the gesture types implemented can be viewed at:

http://www.veoh.com/browse/videos/category/technology/watch/v16618677AhgSDhYE

Although only four gestures were implemented, we postulate that the control algorithm is suitable for any gestures that do not require precise gesture trajectories. From the gesture analysis literature [3][4] it seems clear that beat gestures, deictics and most metaphoric gestures meet this requirement; most iconic gestures on the other hand require precise trajectories in order to have meaning (e.g. outline tracing, movement path description). We plan to investigate what other types of gestures may be effectively produced using our control algorithm. We plan to do this by producing a monologue accompanied by multiple gestures. This will also allow us to assess the effects of gestures on human engagement with the robot.

REFERENCES

[1] T. Fong, I. Nourbakhsh, K. Dautenhahn. "A survey of socially interactive robots." In *Robotics and Autonomous Systems*, 42, pp. 143–166. 2003.

[2] A. J. Cowell and K. M. Stanney. "Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility." *Int. J. Human-Computer Studies,* vol.62, pp. 281–306. 2005.

[3] A. Kendon. *Gesture: Visible action as utterance*. Chap. 6-9. Cambridge University Press. 2004.

[4] D, McNeill. *Hand and mind.* Parts 2 and 3. University of Chicago Press, 1992.

[5] J. Cassell, H. Vilhjalmsson, and T. Bickmore. "BEAT: the Behavior Expression Animation Toolkit." In Proc. *the 28th annual conference on Computer graphics and interactive techniques*, 2001, pp477-486.

[6] M. Stone, D. DeCarlo, I. Oh, C. Rodriguez, A. Stere, A. Lees, and C. Bregler. "Speaking with Hands: Creating Animated Conversational Characters from Recordings of Human Performance". *ACM T. Graphics, 23(3)*. 2004. pp506-513.

[7] M. Neff, M. Kipp, I. Albrecht, and H. Seidel. "Gesture Modeling and Animation Based on a Probabilistic Re-Creation of Speaker Style". *ACM T. Graphics*, 27(1). 2008. pp1-24.

[8] S. Kopp, and I. Wachsmuth, 2004. "Synthesizing multimodal utterances for conversational agents." *J. Computer Animation and Virtual Worlds*, v.15 no.1. 2004, pp. 39-52.

[9] H. Ishiguro, T. Ono, M. Imai1, T. Maeda1, T. Kanda1, and R. Nakatsu. "Robovie: an Interactive Humanoid Robot". *Int. J. Industrial Robot*, Vol. 28, No. 6. 2001

[10] C. Sidner, C. Lee, C. Kidd, N. Lesh, and C. Rich. "Explorations in engagement for humans and robots." *J. Artificial Intelligence*, Vol. 166, Issue 1-2. 2005

[11] Breazeal, C., Kidd, C., Thomaz, A., Hoffman, G., and Berlin M. "Effects of Nonverbal Communication on Efficiency and Robustness in Human-Robot Teamwork". In Proc. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2005.

[12] M. Bennewitz, F. Faber, D. Joho, and S. Behnke. "Fritz - A Humanoid Communication Robot". In Proc. *16th IEEE International Conference on Robot & Human Interactive Communication.* 2007. pp1072-1077.

[13] H. Kim, H. Lee, Y. Kim, , K. Park, and Z. Bien. "Automatic Generation of Conversational Robot Gestures for Human-friendly Steward Robot." In Proc. *16th IEEE International Conference on Robot & Human Interactive Communication,* 2007, pp. 1155-1160.

[14] T. Flash T, and N. Hogan N. "The coordination of arm movements: an experimentally confirmed mathematical model". *J Neuroscience 5*: pp1688-1703. 1985