2011 IEEE-RAS International Conference on Humanoid Robots
Bled, Slovenia, October 26-28, 2011

The Effects of Robot-Performed Co-Verbal Gesture on Listener Behaviour

Paul Bremner
and Anthony G. Pipe
and Chris Melhuish
Bristol Robotics Laboratory
Bristol, UK
Email: paul.bremner
chris.melhuish
tony.pipe
@brl.ac.uk

Mike Fraser
and Sriram Subramanian
Department of Computer Science
University of Bristol,
Bristol, UK
Email: mike.fraser
sriram
@bristol.ac.uk

***Abstract*—Co-verbal gestures, the spontaneous gestures that accompany human speech, form an integral part of human communications; they have been shown to have a variety of beneficial effects on listener behaviour. Therefore, we suggest that a humanoid robot, which aims to communicate effectively with human users, should gesture in a human-like way, and thus engender similar beneficial effects on users. In order to investigate whether robot-performed co-verbal gestures do produce these effects, and are thus worthwhile for a communicative robot, we have conducted two user studies. In the first study we investigated whether users paid attention to our humanoid robot for longer when it performed co-verbal gestures, than when it performed small arm movements unrelated to the speech. Our findings confirmed our expectations, as there was a very significant difference in the length of time that users paid attention between the two conditions. In the second user study we investigated whether gestures performed during speech improved user memory of facts accompanied by gestures and whether they were linked in memory to the speech they accompanied. An observable affect on the speed and certainty of recall was found. We consider these observations of normative responses to the gestures performed, to be an indication of the value of co-verbal gesture for a communicative humanoid robot, and an objective measure of the success of our gesturing method.**

I. Introduction

An obvious means for a humanoid robot to interact with a human is in a natural human-like way, doing so will ideally enable users to engage the mechanisms normally employed in human-human interaction, leading to an intuitive and successful interaction [1]. Co-verbal gestures are the spontaneous gestures that accompany human speech, and have been shown to be an integral part of human-human interactive communications [2][3]. Thus, it seems reasonable to suggest that a humanoid robot should perform co-verbal gestures to interact in the suggested human-like way. Further, it has been demonstrated in anthropological studies that co-verbal gestures have a number of positive effects on listener behaviour [3][4][5]; it is suggested here that gestures performed by a humanoid robot might engender similar effects, and two user studies investigating this idea are presented in this paper.

This observation, of users responding to a robot as they would a person, is described as a *normative response*. Cassell [6] suggests that it represents a form of intersubjectivity, and can be seen as an objective indicator of the efficacy of interactive robot behaviours. Indeed, a range of work has been conducted based on this idea. Sidner et al. [7] demonstrated that mutual entrainment of gaze was observed using a simple humanoid robot, and that it improved user engagement in an interaction; as it had in the human-human interaction studies they conducted. Ono et al. [8] showed that, through correct torso alignment of a direction giving robot, mutual gesturing was observed to occur. Mutlu et al. [9] have demonstrated that human interlocutors respond to gaze cues when performed by a robot as they do human performed cues. Breazeal et al. [10] showed that, by performing human-like affective gestures, their robot's internal state was better understood by users, improving performance on a collaborative task.

A key difference between the related work described above and that presented here, is that the effects tested in the related work are directly focused on the communication of semantic information within the gestures. In this work, on the other hand, the gestures used are an approximation of the spontaneous gestures that typically accompany human speech; thus, their effect on observer behaviour is not directly related to the information they convey. In our previous work, a methodology for producing such co-verbal gestures [11], and rules for improving the human-likeness of sequences of such gestures [12] using our humanoid robot BERTI (Bristol and Elumotion Robotic Torso I), was described. The result of this work was a monologue with an associated script of gestures which we had subjective evidence for the success of. Clearly an objective measure of the efficacy of the gestures is required to fully verify the success of the production method at producing useful gestures and, more importantly, to provide evidence of the usefulness of co-verbal gestures for a communicative humanoid robot. Bennewitz et al. [13] provide some anecdotal evidence that co-verbal gestures improved user interest in their humanoid museum guide robot. Their

robot produces similar sorts of co-verbal gesture to those used here, which, when performed along with other human-like communicative behaviours (e.g. facial gestures), appeared to improve user interest in the robot.

The effects on listeners of co-verbal gestures that are investigated in this paper are those related to listener attention and memory. It is suggested that, in order to maintain listener attention during a speech, the orator should gesture [14]; further, experimental evidence shows that gestures improve observer perception of the quality of a speech and, by extension, their interest in it [15]. Indeed, when studying monologues performed by chat show hosts (professionals at maintaining audience attention) as part of previous work [12], we noted that gestures accompanied the majority of the hosts' speech. Hence, the formation of hypothesis H1: Robot-performed co-verbal gesture will significantly improve the attention span of users. The first user-study investigates hypothesis H1.

Church et al. [4] showed that listeners are able to recall more information of that conveyed to them when the speech is accompanied by gestures than in an audio only condition. Further, elements of the speech that were accompanied by gestures were more likely to be recalled; from this they concluded that gestures are stored differently in memory from speech due to their aid to recall. Additionally, Master et al. [16] suggested that a key indicator of how well something has been remembered is the speed at which facts are recalled; thus, it could be used as an additional indicator (to quantity of data recalled) of the effect of gestures on listener memory. An additional theory on how gesture and its relation to speech is stored in memory was proposed by McNeill [3]. He showed that when participants were asked to repeat a story that had been told to them, they tended to repeat the gestures that accompanied the original story. Hence, we have formed two further hypotheses, H2: Robot-performed co-verbal gesture will significantly improve the recall of facts that are accompanied by gestures; H3: When recalling facts related by a gesturing robot, a person will replicate the gestures performed by the robot. H2 and H3 are investigated in the second user-study.

The user studies that we have conducted utilise our robotic humanoid platform BERTI (Fig. 1), and we have developed work that we have previously conducted [11][12], in order to produce a suitable gesture accompanied monologue for use in the studies; both these elements are described in Section II. The first user study, described in Section III-A, confirms hypothesis H1 that robot-performed co-verbal gesture significantly improves the attention span of users. The second user study, described in Section III-B, confirms that there is a link between robot-performed co-verbal gestures and how well facts are remembered by users (hypotheis H2); this is indicated by a significantly shorter duration of pauses between fact recall, when co-verbal gestures had been performed. However, no evidence was found to support hypothesis H3. From these findings we conclude, in Section IV, that robot-performed co-verbal gestures are able to induce normative responses in users, and are thus a worthwhile behaviour for communicative humanoid robots. Further, these findings provide objective evidence for the efficacy of gestures produced using our proposed simple control scheme. Additionally, we suggest reasons why all of the expected, normative responses were not observed, and hence directions for further investigation.

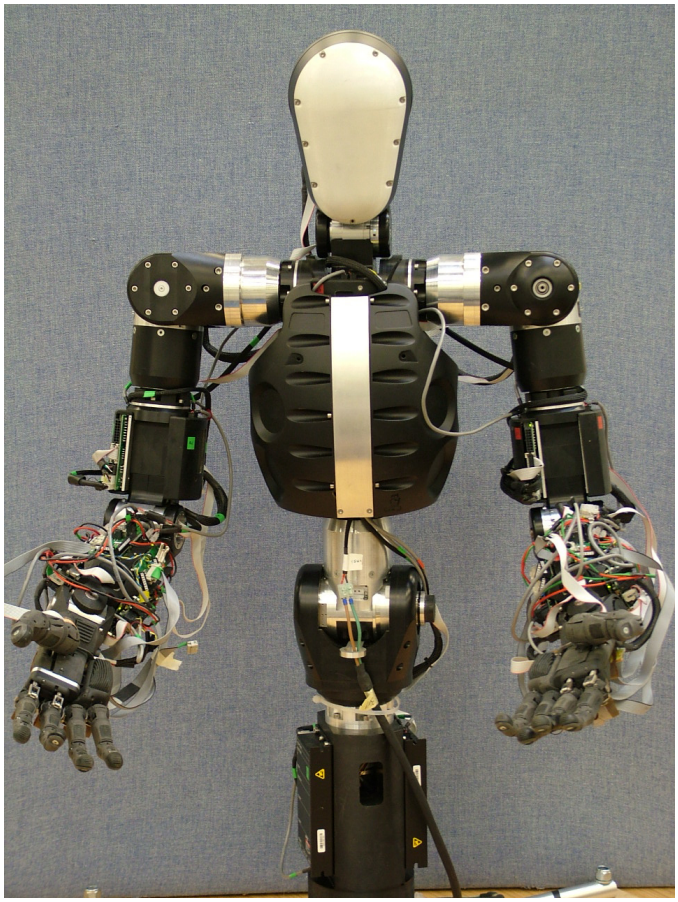

Fig. 1. BERTI, our humanoid robotic platform.

II. Gesturing System

In order to test whether gestures performed by a humanoid robot would engender the expected effects on listeners, we endeavoured to create a method of producing a monologue accompanied by human-like gesture sequences, to be performed by our robotic platform. In the following sections the robot platform is described, along with the methods used to produce, and details of, the required gesture-accompanied monologue.

A. Robotic Platform

In our experiments, we used our bespoke humanoid robot torso BERTI (Bristol and Elumotion Robotic Torso I, Fig. 1). Each arm has seven Degrees of Freedom (DoFs), and there are nine DoFs 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. Each joint has a local motor controller commanded from the base PC using a CAN bus. The arm joints are capable of speeds similar to that of human movement. However, due to mechanical limitations, the

Gesture	Description	Examples of Use
Beat gesture	Rhythmic gesture, with a two movement stroke phase, moving the hand(s) either up and down or in and out	Highlighting important elements of the speech
Indicating a referent	Pointing towards an example of things that are being talked about	Gesturing towards the audience when talking about people, and towards itself when talking about itself or other humanoid robots
Representation of a concept or object	A gestural representation of something that is being talked about	Moving both hands together for two things being joined. Moving hands out from the central gesturing space for 'biggest'. Hands forming a circle when talking about something that is round.

TABLE I
OVERVIEW OF GESTURES PERFORMED BY BERTI DURING THE MONOLOGUE.

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.

The face of BERTI has purposefully been left blank, effectively neutralising facial gestures as a communication channel. We have done this to try to ensure that observed effects are purely attributable to the gestures that are performed. Further, any errors in facial gestures might have had an adverse affect on user perceptions and thus confounded results.

B. Movement Production

In order to produce the movements required for gesture using BERTI, we use a relatively simple control scheme, as we believe it is possible to produce effective gestures using such a method [11]. The inverse kinematics are solved for the end points of each phase of a gesture. The calculation is constrained by specifying the horizontal component of the forearm vector, which can be intuitively specified for the required points necessary to produce gestures. Using triangular joint velocity profiles, the accelerations are calculated so that all the joints will start and finish moving concurrently. This control scheme produces motion that possesses key features of a human-like trajectory (between the two end points), i.e., smooth, direct motion. We have identified these features as key, since they are common to suggested models of human arm motion [17][18]. Our control scheme has been shown to produce demonstrably well rated gestures [11], whilst being significantly simpler to implement than other suggested models.

C. Monologue Production

A monologue lasting approximately 2 minutes, describing some of the research activities in our lab, was written for BERTI to perform in the studies. The text was read using the Microsoft text-to-speech (TTS) engine, prosodic and other paralinguistic information was purposefully left out of the speech, to neutralise it as a communication channel for the same reasons as the blank face given to BERTI; the speech was output through speakers mounted directly behind BERTI's torso. Sequences of gestures were carefully scripted to accompany the monologue to produce the required gestural condition. The gestures that are scripted to accompany the speech are designed to complement the speech, rather than provide semantic content necessary to understand the information being conveyed. A large proportion of the gestures scripted were beat gestures, rhythmic gestures that add prosodic information, typically accompanying elements of the speech that are salient to the speaker [2]. Additionally, gestures were produced that reflected the semantic content of the speech. Examples include, forming a circle with its hands when talking about a round object and indicating the audience when talking about people; a more complete description of the sorts of gestures performed is presented in Table. I. The gestures used in the monologue are similar in form to those tested in our previous work [11], thus we are confident of their means of production.

In anthropological studies, gestures have been described as consisting of three phases: a preparation phase, where the hand(s) move into position; a stroke phase, the most effortful part of the gesture that coincides with the word it is planned to accompany; a retraction phase, where the hand(s) return to rest [2][3]. The preparation phase and retraction phase are optional for a single gesture, dependent on the hand locations preceding and following the stroke phase, i.e., where a gesture falls within a sequence of gestures. Kendon [2] suggests that a sequence of gestures, from when the hands move to the gesture space until they return to rest again (termed an excursion), is typically aligned with an element of discourse which he terms an *idea unit*. In order to gain a concrete understanding of how this might be used to successfully script the gestures for the monologue, videos of chat show hosts performing monologues were examined. The videos were studied taking note of the appearance of the phases of gesture, the timing of phases relative to the speech, elements of speech typically accompanied by gesture, and when gesture units began and ended relative to the speech content. From the study we learnt the form and timing of the different phases, what constitutes an idea unit, and gained an intuitive understanding of when gestures should be performed. Based on what we learnt, the gesture script, including timings and movements, was written. Excursions were scripted to commence just before, and end immediately after, their identified idea units. In addition, the designed monologue was rehearsed by human performers and their gestures observed.

Having determined the desired script for all the gesture

excursions present, the movements needed to be correctly synchronised with the speech content. Stroke duration of gestures was determined by the length of the word they are set to accompany using word times returned from the TTS engine; preparation and retraction phase timing was determined so as to give a human-like movement rate for that motion. Locations within the text, when gesture phases should be executed, are identified and marked using XML tags that can be interpreted by the TTS engine, and thus trigger the initialisation of motion.

III. User Studies

In order to investigate whether co-verbal robotic gestures produce expected beneficial effects on listener behaviour, and provide an objective measure of the success of the described gesture production system (supplementing the subjective evidence previously found [11]), two user studies have been conducted. In the first user study, the effect of gestures on user attention is investigated. In the second study, gesture effects on user recall of conveyed data is investigated.

A. Effect of Robot-Performed Gestures on Listener Attention

It is suggested that, in order to maintain listener attention in a speech, the orator should gesture [14]; further, experimental evidence shows that gestures improve perception of the quality of a speech, and by extension their interest in it [15]. In this context, attention is defined as the listener paying attention to the speaker, i.e., not looking away from the robot for extended periods of time. This user study has been designed to investigate whether users pay attention to a robot-performed speech for longer when it is accompanied by human-like co-verbal gestures.

1) Methodology: In order to assess the attention of listeners, BERTI was set up as part of a departmental display, at an open-day, at the University of Bristol. This venue was chosen for two reasons. Firstly, by setting the robot up in a public space, experimental participants were free to come and go as they pleased and felt no obligation to pay attention to (or indeed stay for) the entire duration of the monologue. Secondly, there was other related activity going on in the nearby area that could draw participant attention if the robot was not sufficiently engaging. The related activity consisted of an autonomous tabletop mobile robot demonstration, and a researcher available to answer questions; it was coordinated with BERTI's performances to ensure as consistent an environment as possible for each performance. The organisation of the open-day meant that, at regular intervals throughout the day, groups of attendees would gather in the area in which the robot was set up. This gathering determined the schedule that was used for the robot's performances, i.e., it was ensured that there were sufficient people in the area (at least 10) before commencement of a performance, to facilitate (as far as possible) a degree of consistency in terms of number of participants for each performance. The reactions of a total of 106 participants (42 female) were analysed in total, participants were made up of a mixture of prospective students and their parents (estimated age range 18-55). BERTI was set up in an area that was separate from the rest of the open day activities so environmental disturbances were minimised. Additionally, during each performance, the environment was monitored for disturbances that might affect the results; several trials were rejected as a consequence of observed disturbances. See Fig. 2 for a diagram of the area in which BERTI was set-up.

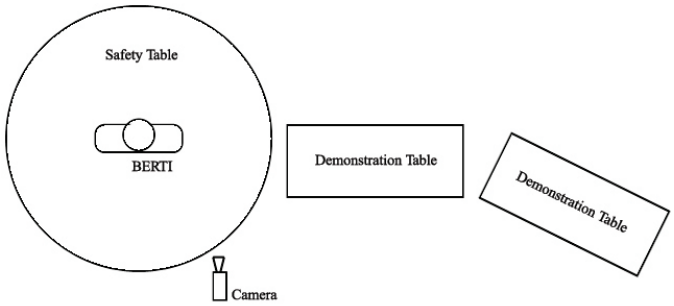

Fig. 2. The experimental set-up used to investigate engagement. Participants were free to move about anywhere in front of the tables, but in order to properly see and hear BERTI they needed to stand in front of the table it was on.

Two conditions were used, one with human-like co-verbal gestures and one without, varied between groups of participants. In the condition without co-verbal gestures, the robot made hand movements in the area in front of its torso (where gestures are typically performed) while it was speaking; the movements were designed to have similar movement characteristics (movement range, velocity etc) as the performed gestures in the other condition, and were performed with similar frequency. This can be considered as a gesture-free (control) condition as the hand movements bear no relation to the speech content; if an unmoving robot was used for this condition, rather than random movement, it would not be possible to determine if it was co-verbal gestures, or simply a moving robot that was holding participants' attention. Each group of participants was only subject to one of the conditions, otherwise the lack of new information in a repetition of the same speech (and hence less interest to a participant) would be likely to confound any results.

Each trial was started when sufficient participants had gathered in the area around the robot. Participants standing in the area in front of BERTI were filmed so that their responses could be analysed afterwards. The fact that filming was taking place was clearly signposted, with signs describing the experiment, and detailing how to opt out if they were filmed against their wishes; this was in accordance with ethical guidelines as laid down by our university's ethics committee. The two gesture conditions were performed sequentially throughout the day, consequently the time stamp on the video segments could be used to identify which condition was used. Three successful trials (i.e., no environmental disturbances) for each condition were performed over the course of the day.

2) Results: The video sequences were analysed by noting, every 30 seconds, the percentage of participants that were

paying attention to the robot, which we define as those that were looking directly at it (based on work attributing gaze in this way [19][20][21]); additional qualitative observations of the data were also noted. At each notation time, 5 seconds of footage was used to determine the percentage of participants paying attention, for example, for notations made on the 30 second interval, footage from 28-32secs (inclusive) was used. At the beginning of a trial the participants present were noted, then during each notation period the proportion of time that each participants' gaze was directed at BERTI was noted, those with a proportion of at least 75% of that period were deemed to be paying attention. Although only a single person video coded the video, we believe the metric is sufficiently unambiguous that the data is reliable. Table II shows the percentage of engaged participants (of those present at the start of the trial) at the fixed time intervals, averaged for the three performances of each gesture condition.

Time (sec)	Co-verbal Gestures	Unrelated Movement
30	97.2	79.3
60	93.9	6.7
90	91.4	3.3
120	91.7	3.3

TABLE II
PERCENTAGE OF VISITORS PAYING ATTENTION TO BERTI AT INTERVALS DURING THE MONOLOGUE.

3) Discussion: A single tailed Chi squared test with Yates correction performed on the data shows a highly significant result ($p < 0.0001$, Chi=39.8, df=1), thus there is a strong relationship between the performance of co-verbal gestures, and audience attention; verifying hypothesis H1. It was noted that the majority of people had not only stopped paying attention to BERTI, but had walked away entirely, within the first 60secs of the speech with random movement. Conversely, when appropriate gestures were performed, the vast majority of people stayed until the end of the performance, and they appeared to be paying full attention to the robot.

B. Effects of Robot-Performed Gestures on Information Recall

Experiments in human-human interaction have shown that portions of speech accompanied by gesture are better recalled by listeners [4]. Another effect on listeners that has been found in anthropological studies, is that gestures performed by speakers are cognitively linked by observers to the associated parts of the speech content they accompany [3]. It is hypothesised here that similar effects should be found in the case of humans observing a humanoid robot performing speech with gestures; i.e, robotic gestures will have similar effects as their human counterparts. This user study has been designed to investigate this hypothesis.

1) Anthropological Theory: In order to properly investigate if a robot gesturing system can produce the effects that have been observed in anthropological studies, it is necessary to understand the experiments that were conducted. The experimental design used in those studies then leads to the design of the user study presented here.

Church et al. [4] showed that speech accompanied by gestures is better recalled than when gestures are absent. In the study that they carried out, participants watched video stimuli of extracts of social conversation, some watched videos with gesture, and some without. When asked to write recollections of the video stimuli, participants who were subject to the gesture-present case were observed to be better able to recall what they had observed. Further, the parts of the speech that were accompanied by gestures were significantly better recalled than those that were not. They suggest that these results indicate that gesture is processed along with speech by listeners, and may have a different status in memory to speech. However, quantity of data recalled was the only indicator used for the ability of participants to remember the data; Master et al. [16] suggest that an additional key indicator of how well data is remembered is the speed of recall. We suggest that both indicators should be used to test the effect of gesture on memory.

Another effect of gestures on listeners was identified by McNeill [3], who proposed the theory that gestures which co-occur with speech are mentally associated (by listeners) with the speech content they accompany. In order to provide evidence for this theory, he performed an experiment whereby a participant was asked to watch a cartoon, who then had to describe it to a second participant; the second participant then re-told the description of the cartoon to a third participant. It was observed that in the retelling of the cartoon description, that the secondary participant often performed similar gestures to those that were used by the first, with the appropriate sections of the description.

2) Experimental Procedure: In order to test for the gesture effects on listeners described above, a user study was conducted where participants would listen to a monologue about work conducted in our lab, and then be asked to recall as much of it as possible. Two different conditions were used, audio-visual and audio-only, varied between subjects. The audio-visual (speech and gestures) condition was similar in form to that used by McNeill [3], however, the participant who gave the original telling of the cartoon description was replaced by BERTI, thus, each participant (in this condition) watched BERTI perform a monologue. In the audio-only (control) condition BERTI was absent, but the same speech content produced by the text to speech engine was used. This control condition was selected as it is closest to the treatments used by Church et al. [4]. Further, in the work of Kawas et al. [22] it was shown that there is no significant difference between the efficacy of memory tests (where data is read to a participant and than asked to be recalled) conducted by phone (equivalent to our control condition) compared to in person (equivalent to a stationary robot with audio).

Prior to being subjected to the stimulus, each participant was informed that they would be asked to recall as much as they were able to of what they heard, to a designated listener, while they were filmed; the retelling was filmed to enable analysis of how much information was recalled, the speed of recall and what gestures were performed by the participant. The listener

was required to provide back-channel feedback, and thus give a more natural circumstance to retelling the data, than might have been achieved by asking the participants to talk directly to the video camera. Safety precautions, data protection and other procedural details were explained to each participant in order to establish informed consent to participate, in accordance with guidelines laid down by our university's ethics committee. 24 participants (11 female) aged 20-35 (M 26.2) took part in the study. Participants were recruited from outside the robotics laboratory, were English, and had minimal prior experience with both robotics and gesture analysis.

For the gesture performances, the above described informational monologue was used. Although the thematic content of the speech differed from that used in the experiments of McNeill [3] and Church et al. [4], there was sufficient gestural content that we expected gesture effects, resulting from robot gesturing as hypothesised, should be observed.

3) Results:

a) Information Recalled: In order to analyse the speed of recall, and the quantity of data recalled by each participant, the monologue was broken down into 21 elements that might be recalled. Of these elements 10 were accompanied by beat gestures, 8 by non-beat gestures, and 3 were unaccompanied. The video of each participant's retelling was than examined, noting the number of these data elements that were recalled, and the duration of pauses between each element. The mean duration of total pauses is shown in Fig. 3, and the mean number of elements recalled in the two conditions is shown in Fig. 4. Although only a single person coded the video, we believe the metrics are sufficiently unambiguous that the data is reliable.

An unpaired 2-tail t-test performed on the data showed that the pauses were significantly longer in the audio only condition (df = 22, t=2.31, $p < 0.05$). Further, Master et al. [16] also suggest that affective tone is an indicator of how well a piece of data has been remembered. Qualitative analysis of the participants' responses showed (in the authors' opinion) that there was a marked difference in affective tone between the two conditions; participants in the audio-only condition were audibly less certain. It was also noted that 42% of the audio-only participants corrected elements of data that they initially recalled incorrectly; no corrections were observed in the audio-visual condition. Recall of the elements unaccompanied by gestures did not appear to vary between the two conditions, and there was no consistent difference in effect on recall between beat and non-beat gestures across participants.

We suggest that the significant difference in pauses in data recall, and supplementary qualitative analysis, suggests that the confidence in the information recalled is improved by the performance of co-verbal gestures by BERTI; these findings support hypothesis H2.

However, an unpaired 2-tail t-test performed on the data showed there was no significant difference between the two conditions in the quantity of information recalled (df = 22, t=0.75, $p = 0.46$). This is contrary to the findings of Church et al [4], and hence our expectations. Reasons for this finding are suggested in the following discussion section.

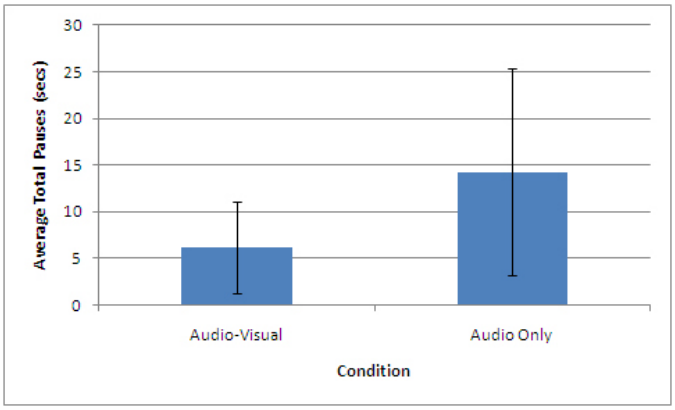

Fig. 3. Total duration of pauses during retelling

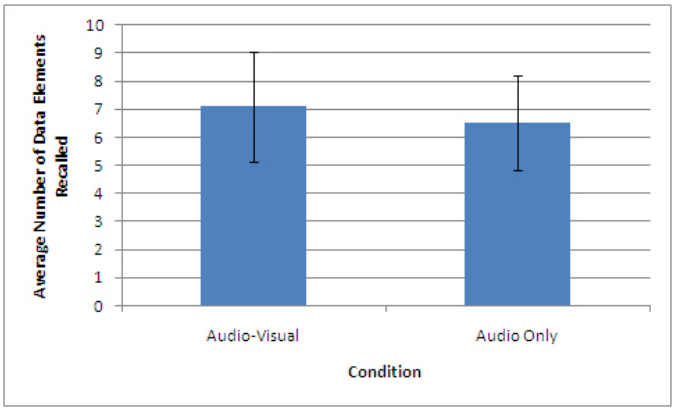

Fig. 4. Number of elements of data recalled

b) Gestures Performed By Participants: Participants in the audio-visual condition were not, in the majority of cases, observed to repeat any of the gestures performed by BERTI; two of the twelve participants performed one of the gestures. This is contrary to the findings of McNeill [3], and hence we reject hypothesis H3. Reasons for this finding are also suggested in the following discussion section.

4) Discussion: It was clear from analysis of the videos that the gestures performed by BERTI had an effect on participants confidence in the information they recalled, although not on the quantity of data recalled. We plan to perform a more detailed analysis of the video data to better understand this effect, which has important implications for future work on human-robot interaction. If the data conveyed is of an instructional nature, significantly longer pauses in recall of instructions, and uncertainty of the data recalled, would be likely to have a significant impact on task performance. A possible explanation for the different manifestation (between the work of Church et al. [4] and that presented here) of the gestures' effect on memory, is the variation in method by which participants were asked to recall the data. In the work of Church et al. [4] participants were asked to write down what they could remember, so analysis of the confidence of

participants in the data they recalled was not possible; a factor we felt was important to investigate. Thus, the use of oral recall was designed to be more instructive than written recall would have been. However, this difference in recall methods also suggests one possible explanation that no difference in the quantity of data recalled was observed here, it has been shown that data is better recalled when performed vocally (compared to written recall) [23]. However, despite the difference in form, an effect on memory is clearly observed; thus suggesting that robotic gesture, like human gesture, is stored differently in memory to speech.

Although almost no evidence was found of gestures performed by BERTI being performed when participants repeated the information, there are three possible explanations which merit further investigation. One possibility is the difference in the purpose of the gestures between those investigated here and those in McNeill's experiment. In the experiment described by McNeill [3] the monologue to be retold was a description of a cartoon, thus gestures performed often conveyed some of the detail of the events that occurred. Whereas, in the work presented here, the gestures investigated did not convey any additional semantic information, and were thus only required to convey which elements of the speech were salient. This might suggest that this conveyance of additional information by the gestures is a requirement for their reproduction during retelling. An alternative explanation is that, in order to engender the required degree of normative behaviour to observe this effect, a more human-like conveyance of information is required; i.e., human-like elements not currently implemented, such as better artificial speech and a robotic face. Finally, the fact that the contents of the speech were known to the listener, a fact obvious to the retellers, may have influenced the gestures performed; gestures are more frequently seen accompanying speech content not currently in the shared knowledge space of speaker and listener [24].

IV. Conclusion

In order to investigate whether robotic co-verbal gesture has similar effects (on listeners) to human co-verbal gesture, and provide objective evidence for the efficacy of the gestures generated, two user studies have been conducted. The first study showed that people paid attention to BERTI for significantly longer when it performed co-verbal gestures, than when it moved in a way unrelated to the speech content; this matches the suggestion that human orators who gesture are better able to hold audience attention [15][14], i.e., confirming hypothesis H1. This was done by BERTI performing a monologue both with, and without, co-verbal gestures in a location where participants were free to remain and pay attention, or leave as they desired.

In the second study it was shown that the gestures performed by BERTI had an effect on the confidence participants has in recalled information. When gestures were performed there was a significantly shorter duration of pauses between elements recalled, than in the audio only condition. Supplementary to this, participants in the gesture condition had a more confident affective tone (in the authors opinion), and they did not have to correct statements they had made; corrections were observed in the recall of several of the audio-only participants. This shows a clear link between gestures and information recall. Although the effect observed is not identical to that of Church et al. [4], as a consequence of the work of Master et al. [16] we regard it as similar in its implications for the role of gesture in memory, and strong evidence in support of hypothesis H2.

The observation of these effects produced by robotic gesture shows that gestures produced using simple heuristics are able to induce somewhat normative responses from people, providing objective evidence for the efficacy of our method of gesture production. We suggest this has significant implications for future work on behavioural design of communicative humanoid robots. We propose that such robots should be endowed with the capability to produce co-verbal gestures in order to improve the efficacy of their communications. Further, while the movements and behaviours must, to some degree, be based on a human model, they need not be a perfect reproduction; the degree of human-likeness, and accuracy of synchronisation required merits further study.

A caveat to our findings is that not all of the effects sought were observed. In the second user study the validity of hypothesis H3 was investigated, i.e., whether the robotic gestures would be reproduced by participants during recall; as was observed in human studies by McNeill [3]. Although this effect was not observed, some explanations that merit further investigation were suggested. Firstly, the content of the speech and the types of gestures performed (whether human or robot performed) needs to be investigated as to the identify the important characteristics necessary in order for them to be repeated during retelling. Secondly, BERTI may need to be more human-like in order to induce the effect; for example, better artificial speech and a robotic face. Additionally, Church et al. [4] observed an effect on the quantity of data recalled, for which no significant evidence was found with BERTI. Our investigation of the confidence in data recalled may have been a reason for this, i.e., data recall has been observed to be easier when performed vocally than when written [23]. Thus, our expectations for the observation of effects of robotic gestures on listener behaviour have been partially confirmed, and we are motivated to conduct further work investigating them.

A. Future Work

One of the noted possible reasons that all expected normative responses to the gestures were not observed is that all conversational communication channels might be required in order to induce them to occur, i.e., gestures may not be the solely responsible factor. Thus, a possible direction for future work is to develop a more complete, more human-like communication system. Two identified key areas for doing this are facial gestures, and paralinguistic information in the generated speech. Implementation of both of these modalities would need to be carefully considered, as it is important that they are performed correctly, and act in synchrony with

the produced gestures; errors in one communication modality seem likely to occlude the benefits of another.

Another possible line of investigation is to alter the content of the speech, and thus the script of gestures that is used for investigation of the described effects. It seems possible that different types of information (factual, conversational etc.), and different types of gestures, will be remembered differently, and thus influence how gesture effects on memory are manifested.

REFERENCES

[1] C. C. Kemp, P. M. Fitzpatrick, H. Hirukawa, K. Yokoi, K. Harada, and Y. Matsumoto, "Humanoids," in *Springer Handbook of Robotics*, 2008, pp. 1307–1333.

[2] A. Kendon, *Gesture: Visible Action as Utterance*. Cambridge, UK: Cambridge University Press, 2004.

[3] D. McNeill, *Hand and Mind: What Gestures Reveal About Thought*. Chicago, USA: University of Chicago Press, 1992.

[4] R. B. Church, P. Garber, and K. Rogalski, "The role of gesture in memory and social communication," *Gesture*, vol. 7, no. 2, pp. 137–158, August 2007.

[5] S. Goldin-Meadow, "The role of gesture in communication and thinking," *Trends in Cognitive Sciences*, vol. 3, no. 11, pp. 419–429, November 1999.

[6] J. Cassell and A. Tartaro, "Intersubjectivity in human-agent interaction." *Interaction Studies*, vol. 8 (3), pp. 391–410, 2008.

[7] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," *Artificial Intelligence*, vol. 166, 2005.

[8] T. Ono, M. Imain, and H. Ishiguro, "A model of embodied communications with gestures between humans and robots," in *Proceedings of the 23rd meeting of the Cognitive Science Society*, 2001, pp. 760–765.

[9] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita, "Footing in human-robot conversations: how robots might shape participant roles using gaze cues," in *HRI '09: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, 2009, pp. 61–68.

[10] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin, "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005, pp. 383–388.

[11] P. Bremner, A. G. Pipe, C. Melhuish, M. Fraser, and S. Subramanian, "Conversational gestures in human-robot interaction," in *SMC'09: Proceedings of the 2009 IEEE international conference on Systems, Man and Cybernetics*, 2009, pp. 1645–1649.

[12] P. Bremner, A. G. Pipe, M. Fraser, S. Subramanian, and C. Melhuish, "Beat gesture generation rules for human-robot interaction," in *RO-MAN '09: Proceedings of The 18th IEEE International Symposium on Robot and Human Interactive Communication.*, 2009, pp. 1029 –1034.

[13] M. Bennewitz, F. Faber, D. Joho, and S. Behnke, "Fritz – a humanoid communication robot," in *RO-MAN '07: Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication*, 2007, pp. 1072–1077.

[14] S. Mandel, *Effective presentation skills*. London, UK: Kogan Page, 1987.

[15] S. B. Fawcett and K. Miller L, "Training public-speaking behavior: an experimental analysis and social validation," *Journal of Applied Behavior Analysis*, vol. 8(2), p. 125135, 1975.

[16] D. Master, W. A. Lishman, and A. Smith, "Speed of recall in relation to affective tone and intensity of experience," *Psychological Medicine*, vol. 13, no. 02, pp. 325–331, 1983.

[17] T. Flash and N. Hogan, "The coordination of arm movements: An experimentally confirmed mathematical model," *Journal of Neuroscience*, vol. 5, pp. 1688–1703, 1985.

[18] E. Nakano, H. Imamizu, R. Osu, Y. Uno, H. Gomi, T. Yoshioka, and M. Kawato, "Quantitative examinations of internal representations for arm trajectory planning: Minimum commanded torque change model," *Journal of Neurophysiology*, vol. 81, pp. 2140–2155, 1999.

[19] R. Fang, J. Y. Chai, and F. Ferreira, "Between linguistic attention and gaze fixations inmultimodal conversational interfaces," in *Proceedings of the 2009 international conference on Multimodal interfaces*, 2009, pp. 143–150.

[20] Y. I. Nakano and R. Ishii, "Estimating user's engagement from eye-gaze behaviors in human-agent conversations," in *Proceedings of the 15th international conference on Intelligent user interfaces*, pp. 139–148.

[21] C. Rich, B. Ponsleur, A. Holroyd, and C. L. Sidner, "Recognizing engagement in human-robot interaction," in *Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction*, 2010, pp. 375–382.

[22] C. Kawas, H. Karagiozis, L. Resau, M. Corrada, and R. Brookmeyer, "Reliability of the blessed telephone information-memory-concentration test," *Journal of Geriatric Psychiatry and Neurology*, vol. 8, no. 4, pp. 238–242, 1995.

[23] E. Loveman, J. C. van Hooff, and A. Gale, "A systematic investigation of same and cross modality priming using written and spoken responses," *Memory*, vol. 10, pp. 267–276(10), 2002.

[24] M. Halliday, *Intonation and grammar in British English*. Paris, France: Mouton, 1967.