Real-time Teleoperation of an Industrial Robotic Arm Through Human Arm Movement Imitation

Mavridis N., Machado E., Giakoumidis N., Batalas N., Shebli I., Ameri E., Neyadi F., Neyadi A.
Interactive Robots and Media Lab
United Arab Emirates University
Al Ain, United Arab Emirates
irmluaeu@gmail.com

***Abstract*—Real-time remote teleoperation of robotic arms, either industrial or humanoid, is highly desirable for a number of applications, especially in difficult or inaccessible environments. Here, we present a system for teleoperation of an industrial arm commanded through human-arm motion capture, as well as a bi-partite evaluation. Usually, teleoperation is driven by buttons, joysticks, haptic controllers, or slave-arms set, necessitating further training. In contrast, in our system the desired trajectory of the arm is easily and naturally controlled through imitation of simple movements of the operator's physical arm, obtained through motion capture. Furthermore, we present an extensive evaluation of the performance of our system, containing: first, task-related measurables for a fixed task performed by numerous previously untrained subjects; and second, user-opinion/attitude data obtained through a questionnaire administered to experimental subjects. Aspects of the system as well as results are discussed, and future extensions presented.**

I. INTRODUCTION

Numerous application domains of robotics make the physical co-presence of human operators nearby the robot difficult, for example, hazardous or radioactive environments, space, etc. Furthermore, towards full-body android telepresence, teleoperation is implicated as one of the key supporting technologies. Quite some research on teleoperation has take place [1], but most systems rely on unnatural controllers, such as joysticks, which require previous training. Notable exceptions do exist, for example a demonstration of the benefits of using human natural arm movement for controlling an excavator [2]. In that paper, the authors claim to solve two problems usually related with excavators: high risk involved in the operation, and difficulty inherent of manipulation by joysticks. The authors use a combination of orientation sensor, rotary encoder, and inclinometer to read the human arm and hand movement and transmit the data to a computer through bluetooth, which then controls the excavator. In another related work [3], the authors used optical motion caption to copy the operator's arm and head movement to an android. Their intention was to create a natural human-like movement on the android as a way of improving the interaction between it and humans. In our system, we use real-time motion capture, for easy and intuitive teleoperation of an industrial arm, by simple imitation of human arms movements towards completing pre-specified tasks. Regarding the important problem of correspondence choice between imitator and imitated (robot and human in our case), the reader is primarily referred to the extensive analysis in [4], and also to [5]. Our system comprises the following three steps (see Fig. 1): first, the operator's arms movements are captured using optical motion capture; second, we choose an appropriate correspondence, and we apply geometric transformations to the data received from the motion capture, sequence them, and translate them to the robot command language; third, we send the data to the industrial robotic arm. Furthermore, user feedback is provided through two visual and one auditory channel. Apart from the naturalness and the choice of the controllers, another important aspect of relevant research is the evaluation of the complete system. Although time-delay as well as limited spatial aspects of the performance of such systems has been reported, to the best of our knowledge, no task-based evaluation of the reaction of unskilled people during teleoperation of an industrial robotic arm has yet taken place. Therefore, an extensive evaluation also including task-based as well as user-satisfaction and attitude data is presented here, as we discuss below.

Figure 1. Teleoperation System Block Diagram

Proceedings of the 2010 International Symposium on Robotics and Intelligent Sensors (IRIS2010)
Nagoya University, Nagoya, Japan, March 8-11, 2010. ISBN: 987-4-9905048-0-9, ISSN: 1884-1023

II. The System

The tele-operation system consists of five major subsystems: Motion Capture, CyberGlove, TeleOp Controller, Robotic Arm, and User Feedback (Fig. 1).

A. Motion Capture Susbsystem

The motion capture subsystem, consists of cameras operating at VGA resolution (640 x 480) supporting up to 200fps (Standard Deviation brand). The cameras have infrared LED rings around them, and are placed at a height of 2.62m on the corners as well as the short-side midpoints of a rectangle with size 6m by 4.80m. The effective capture area thus has a footprint of roughly 3m diameter. The human is wearing a special suit on which 19 reflective ball markers of diameter 2.5cm are placed. Three types of suits were used (Fig. 2): either strap-based for western-dressed humans, white traditional emirati dresses for men, and black traditional emirati dresses for women. The software API of the mocap system exposes a number of methods in C++, which enable the quasi-realtime readout of the 3D positions of the tracked markers.

B. CyberGlove Subsystem

The 5DT Ultra Series 14 gloves are used, which can produce 14 finger measurements plus 2 accelerometer readings. The gloves provide triggers for controlling the gripper of the robotic arm.

C. TeleOperation Controller Subsystem

The controller reads out the marker position timeseries from the motion capture, performs coordinate mapping and correspondence (Fig. 3), checks limits of movement, and issues appropriate commands to the robotic arm.

Correspondence choice and Coordinate mapping: We have limited our choice of what the robotic arm should imitate to a simple hand position on a cartesian space. Therefore, our system captures the human subject's right hand position relative to his own arm position and map it to the robotic arm's hand coordinates. The robotic arm controller takes care of doing the inverse kinematics and moving its motors in a way its hand goes to the requested position. Two markers on the subject are necessary to perform this operation: the right wrist marker (RW) and the right shoulder marker (RS). The 3D coordinates of the subject's hand are computed by subtracting the coordinates of the RW marker from the ones of the RS marker. As the robotic arm and the human arm are not of the same size, and even different among various subjects, we empirically computed a scale factor in order to match the subjects' fully extended arm to the robot' fully extended arm. Therefore, the coordinates acquired from the motion capture system are multiplied by this factor before being sent to the robot.

The robotic arm also has the ability of moving its hand relative to its wrist. We have chosen to map this movement to the subject left arm. Therefore, we have made a correspondence between the subject's left elbow angle to the direction of the robotic arm's hand. When the subject's left arm is fully extended, the robotic arm's hand points downward and when it is fully contracted, the robotic arm's hand points upward. To compute this angle, we used three markers: left wrist (LW), left elbow (LE), and left shoulder (LS). These markers can be seen as a triangle and, thus, the elbow marker can be easily computed using the cosine rule.

Limits of movement: Due to security reasons, we have limited the robotic arm's movement to 180 degrees on the X coordinate. This means that the subject can move its arm to points where the robot can't, but the robotic arm only goes where it would not crash into its surrounding objects and break itself. Furthermore, the robot’s gripper is not allowed to go below the floor level, up to a safety margin.

Temporal and Software aspects: Another limitation of the robotic arm is on the data rate it can receive. Its controller ignores commands sent when it is performing a movement and thus we implemented a synchronized communication between it and the teleoperation controller. The tele-operation controller software was developed in Java 6 and integrated with the C++ API of the mocap subsystem, using Java Native Interface (JNI).

D. Robotic Arm Subsystem

The ST Robotics ST 17 manipulator arm is used, which has 5 degrees of freedom on the body, plus one for the gripper. Communication to the robot is achieved through a virtual serial port fed by IP, in the form of RoboForth messages. The workspace of the robot is contained within a hemisphere of one meter radius.

E. User Feedback Subsystem

User Feedback is provided through three channels: two visual, and one auditory. The visual channels are video feeds from two cameras placed in the robot's location: one on the gripper, and one on a tripod behind the robot, overlooking it from an angle. The video feeds are shown on two 40" LCD screens in the room of the operator. Auditory feedback is provided through a microphone in the robot location driving a speaker system in the operator location. The feeds are delivered through proprietary camera software, and the VLC player has also been used in the past.

Figure 2. MoCap suits with markers: Western (L), Emirati Women (M), Emirati Men (R). Notice the CyberGlove on the left hand in L

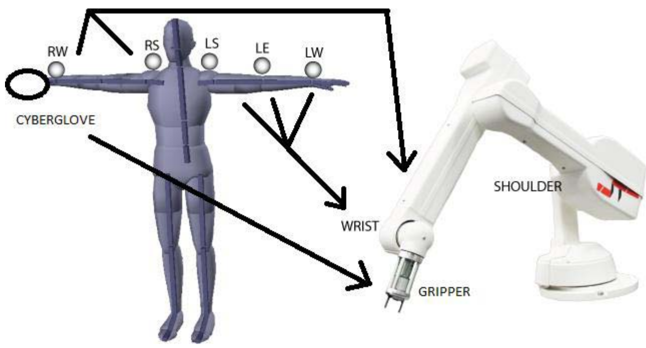

Figure 3. Correspondence choice: Human R hand controls robot position up to wrist, R palm controls gripper, L hand controls robot wrist

III. TASK-BASED EVALUATION

The purpose of our evaluation was to provide real-world experience to our experimental subjects in tele-operation, to investigate the effectiveness of the design choices for our system, and also to create a task transcription and modeling framework which can be used to provide a firm basis for investigating effect of design choices, user variability, as well as aspects of user adaptation and fatigue.

A. The Task

The task chosen was to move three balls from their fixed home positions to their target positions. The task was designed so that it had *intermediate difficulty*, so that we can get meaningful results, without being neither impossible nor trivial. The layout of the workspace for the task is shown in (Fig. 4). The home positions had a 1.5cm elevation from the floor, while the target positions had a paper underneath them with concentric circles marked with 1cm-10cm signs corresponding to the accuracy of the placement.

Figure 4. Task Setup: Robot, balls in home positions, and target markers

B. Administering the Task

The subjects were first exposed to a 5-minute *introduction* of system usage by the person who was administering the task (competent user). Then, the goal of their trial was made explicit: to try to move all the balls to the targets with maximum placement accuracy, as fast as possible, but within 10 minutes. An explicit scoring function was also given to them, in order to remove the arbitrariness of subjective weighing of the *two components* of the goal: first, the number of balls successfully placed (n), taking into account the accuracy of placements (a1, a2, a3) in cm and second, the total time (t) in minutes. I.e. the subjects were told that they had a maximum of 60 points, out of which 30-(t*3) points for total time, and 10*(10-a) points for the accuracy of each ball (which would default to 0 if the ball was not successfully placed). The main purpose of this function was to direct equal importance to both components of the goal for the subjects, so that they don't concentrate more on one of the components, and thus introduce bias. After the goal was made explicit, the subjects were given *5 minutes to play* with the system, and then their up to *trial time* started (maximum allowed duration *10 minutes*), during which video recordings as well as system log files were kept.

C. Transcription and Modelling

Each trial was analyzed on the basis of *six different types of events*: Start, Unsuccessful Grip (UG), Successful Grip (SG), Unsuccessful Placement (UP), Successful Placement (SP), and End. Each trial was thus transcribed as a sequence starting with a Start event, containing a number of UG, SG, UP and SP, and finishing with End. These event sequences were also augmented with the *time intervals* between the events. Transcription was done by humans on the basis of the video recordings. The chosen *underlying model* for the observed data was a probabilistic automaton with 6 states, corresponding to the 6 events. The transition probabilities as well as the transition durations for this automaton were thus estimated on the basis of the observed data, as we shall see.

D. Repetitive Trials

While most of our subjects only had one trial on our system, we chose to perform repetitive trials for a subset of our subjects in order to start investigating learning and fatigue effects, as we shall see in the results section.

IV. OPINIONS AND ATTITUDES EVALUATION

The purpose of this evaluation was to: a) illuminate opinions and attitudes towards the use of tele-operation in different application domains, b) assess the estimated emotional reaction of people in the subject's social circle towards the system, c) to see whether the system demo stimulated subjects to learn more about robotics and tele-operation, d) to gather comments for system improvements.

A. The Questionnaire

The questionnaire had the form of a single-sided A4 sheet, and was available in two languages: Arabic and English, which the subjects could choose. It was partitioned in five parts: demographic questions, opinions and attitudes towards applications, estimated emotional responses, wanting to learn more, and suggestions / comments [6]. The demographic questions queried country of birth, age, sex, college education. Regarding a) and b), a 4-point modified likert scale was used (forced choice), with strongly disagree (1), slightly disagree (2), slightly agree (3) and strongly agree (4) boxes. Regarding a) the seven application areas queried were: medical, workplace, child instruction, games, communication with people, dangerous environments, and space. Regarding b) the four emotional responses queried were happy, comfortable, angry and afraid.

B. *Administering the Questionnaire*

The subjects were given the questionnaire in our lab after going through a standard five-minute introduction to tele-operation, during which a video of our system was shown, as well as a video of android teleoperation, and the benefits of the technology were explained. Most of the subjects that completed the questionnaire also tried out the system themselves.

V. RESULTS

Demographics as well as results for the task-based and the questionnaire evaluation are presented in this section.

A. *Demographics*

29 subjects completed the questionnaire, out of which 23 also tried out the system themselves. Of the 29 subjects, 18 (62%) chose to complete the questionnaire *in Arabic*, and 11 (38%) *in English*. 24 of the 29 subjects, 24 (82%) were UAE nationals, 2 Iranians, as well as 1 Palestinian, 1 Greek, and 1 citizen of the USA. Their *age* ranged between 17…43, while 23 out of the 29 subjects were UAEU students aged between 17…22 years old. The *female to male* ratio was 12:17, i.e. approximately 3:4.

B. *Attitudes towards Applications of Teleoperation*

The results of the attitudes towards the seven application areas are presented in (Fig. 5). One can observe that: (note that here, by agree we refer to the sum of slightly and strongly agree, and by disagree to the sum of slightly and strongly disagree, rounding to 1%)

Hospital: 66% agree, Most: slight agree

Workplace: 83% agree, Most: strong agree

Child Instruct: 31% strong dis, 52% slight agr (bimodal)

Games: 93% agree, Most: strong agree

Comm w Human: 79% agree, Most: slight agree

Dangerous: 93% agree, Most: strong agree

Space: 100% agree, Most: strong agree

One can conjecture the following *preference ordering for teleoperation applications* (order of decreasing preference):

Strong Agree: Space, Dangerous environments, Games

Slight Agree: Workplace, Remote comms, Hospital

Bimodal Slight Agree/Strong Disagree: Child Instruction

After a quick investigation, it was found that the sex (male/female) or age group (17…22 vs. 23…) of the subjects could not predict the two categories (slight agree or strong disagree) apparent in the bimodality of the attitudes towards the use of robots for child instruction.

C. *Estimated Emotions of Peers*

The estimated emotions of peers questions were querying four descriptors of affective states: happy, comfortable, angry, and afraid. The first two have positive valence, the second two negative. From the results in (Fig. 5), one can see that (see comments of subsection above):

Happy: 97% agree, Most: strong agree

Comfort: 72% agree, Most: slight agree

Angry: 86% disagree, Most: strong disagree

Afraid: 83% disagree, Most: strong disagree

Thus, one can conjecture that subjects estimate that their peers (belonging to their social circle) would generally feel happy if they saw the demo. However, the subjects would only slightly agree that their peers would feel comfortable. In contrast, the subjects estimate that their peers would generally not feel angry or afraid.

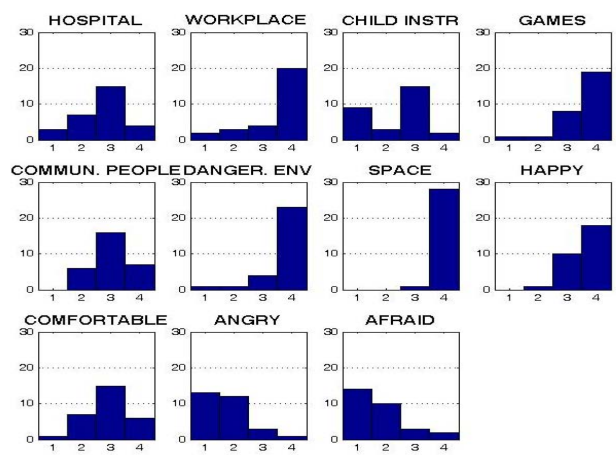

Figure 5. Questionnaire results in histogram form for the seven application domains, and the four estimated emotions of peers

D. *Willingness to learn more*

Twenty five out of 29 subjects answered the two ―willingness to learn more after demo" questions. All 25 (100%) answered positive to the question whether they wanted to learn more *about robotics*, while 2 out of 25 answered No regarding whether they wanted to learn more *about teleoperation*, and 23 out of 25 (92%) answered Yes.

E. *Suggestions and Comments*

Nine out of 29 subjects provided suggestions and comments, 6 of which contained suggestions, regarding *speed, smoothness, delay, and size*: ―Make it more smooth at move", ―I think it be smaller to be easy to use", ―Shorter delay more accurate movements" etc., and 3 of which were congratulatory: ―It was great, I like it very much" etc.

F. Task-based Evaluation: Overall Metrics

As mentioned above, the task was chosen in order to have intermediate difficulty, situated between the trivial and the impossible. For example, our gripper design and the soft balls used often resulted in unsuccessful grips, if the grip position was not precise enough. Across 69 trials, we evaluated the following overall metrics:

Task time per ball (first event to last event):

Median 16sec, Mean 22.7sec, Std 20.7sec

Task final states:

Success (SP) 60.8%, Fail 39.2% (UP 11.6%, UG 27.6%)

Accuracy for successful placements:

Median 6cm, Mean 5.5cm, Std 2.9cm

Number of states per trial:

Median 2, Mean 2.73, Std 1.56

Ratio of overall number successful to unsuccessful events:

SG:UG = 0.92, SP:UP = 2.27 (UGs are often repetitive)

Ratio of probability of Success/Failure of Grip:

P(SG|Start):P(UG|Start) = 1.32

Ratio of probability of Success/Failure of Placement:

P(SP|SG):P(UG|SG) = 2.33

Accuracy Component of Score (0 for UP, 10 for 0cm Acc):

Median 8, Mean 7.8, Std 5.6 (Sum across three balls)

Time Component of Score ((600-Ttotsec)/600*30):

Median 23.2, Mean 22.0, Std 3.46

Total Score (accuracy + time components):

Median 29.85, Mean 29.8, Std 7.37

G. Task-based Evaluation: Derived Model

The probabilistic finite state machine that was derived from our observations can be seen in Fig. 6. Transition probabilities were calculated from the transition matrix resulting from our observations (69 = 3 balls x 23 subjects). Histograms of the transition time distributions are in Fig. 7.

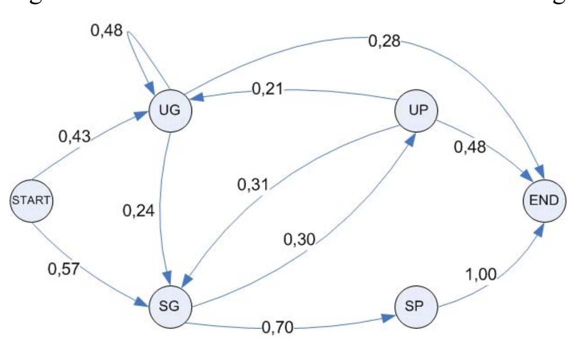

Figure 6. Probabilistic Finite State Machine Model of Task

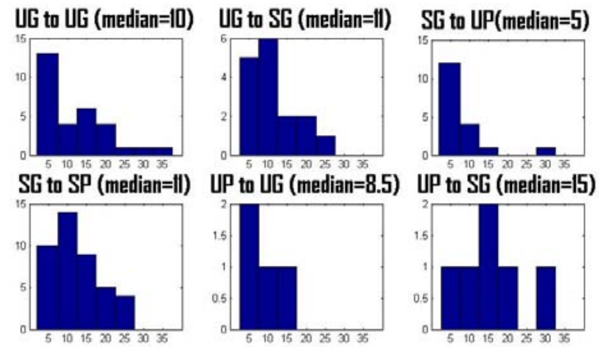

Figure 7. Time interval distributions for state transitions

H. Observations on Derived Model

The derived model, which can be packaged in the form of a transition matrix T, together with the six transition time distributions P(Δt|Si,Sj), and the placement accuracy distribution P(r), provides for a compact description of the performance of the system across users for a single trial, and overall metrics can generally be derived by it. Various observations follow directly: first, according to the score distribution, indeed we have a task which is neither trivial nor impossible; median and mean scores are very near the midpoint of the scale, with a decent amount of variance. Second, upon further analysis, lots of interesting patterns exist in the data: for example, have a look at Fig. 7: Following a successful grip (SG), there are two possible next events - a successful placement (SP) and an unsuccessful placement (UP). The time interval between SG and the next event though is a pretty good predictor of whether it will be successful or not: intervals above 7.5 sec most often lead to success – and whoever rushes often fails - as the time distributions of SG to UP vs. SG to SP seem to indicate. More such patterns remain to be explored, and can be quantitatively supported using probabilistic argument, given more empirical data. The most important observation though has to do with the possible semantics of the {T, P(Δt|Si,Sj), P(r)} description given alternative experimental settings. One can thus ask: how can one try to deconvolve the effects of correspondence choice, user feedback channel, operator ability, learning, and fatigue through such models?

I. Toward insights on Fatigue and Learning

Towards investigating the previous question, five out of the 23 first trial subjects were chosen to continue upon a longer-term study. First, each went through a second prolonged nine-ball session, four days after the first trial. Subsequent sessions are planned, spaced out in time, and with varying lengths. Initial observations indicate a learning effect across sessions, summarized by increasing scores across sessions, as well as a fatigue effect for heavily prolonged sessions, summarized by a score decrease on later trials within a long session. Apart from the score summarization, specifics on the nature of the increase or deterioration, as well as quantitative dependencies, await for more data in order to provide empirical support.

VI. Future Extensions and Discussion

Many possible avenues for future extensions exist. Currently, we are pursuing an extension of the population taking part in our evaluations, and mainly the longer-term multi-session multi-trial evaluations towards insights on fatigue and learning described above. Another avenue that we plan to pursue is concerned with the investigation of the effectiveness of our design choices regarding human- to robot- correspondence. Initially, we had experimented with a glove-less system, in which a different degree of freedom of the human left hand was utilized for controlling the gripper. However, this was found to be highly confusing and difficult to learn for pilot subjects. Still, it is not clear that the current correspondence choice is by any means optimal; so further choices could be potentially investigated. Furthermore, it was noticed that the current two-camera setting for user feedback often does not provide an accurate perception of depth when approaching the ball, which results in misestimation and grip failures. Thus, we plan to investigate alternative camera placements for better results.

Also, many interesting possibilities for pattern recognition and prediction problems based on our task model exist: for example, one could try to predict the score of an individual on the basis of the first 10 seconds or the first ball of his trial; and one could even try to investigate if recognition of an individual through his task-signature is possible, for suitably modified tasks.

Finally, yet another direction which we have started pursuing is migrating our teleoperation system to our conversational Arabic-speaking android robot Ibn Sina [7][8]; in which case it will be used for motion training as well as embodied robotic telepresence, and will cover two hands as well as facial expression imitation. Ibn Sina is part of an interactive theatre, in which various modes of tele-participation are supported, including human-robot interaction through avatars in virtual worlds, remote brain-computer interfacing teleoperation etc.

VII. Conclusion

In this paper a system for teleoperation of an industrial arm commanded through human-arm motion capture was presented, as well as a bi-partite evaluation. Real-time remote teleoperation of robotic arms, either industrial or humanoid, is highly desirable for a number of applications, especially in difficult or inaccessible environments. Usually, teleoperation is driven by buttons, joysticks, haptic controllers, or slave-arms set, necessitating further training. In contrast, in our system the desired trajectory of the arm is easily and naturally controlled through imitation of simple movements of the operator's physical arm, obtained through motion capture. Apart from a detailed description of our system and the design choice made, we presented an extensive evaluation of the performance of our system, containing: first, task-related measurables for a fixed task performed by numerous previously untrained subjects; and second, user-opinion/attitude data obtained through a questionnaire administered to experimental subjects. During the task-based evaluation, which was tuned in order to be neither trivial nor impossible, a probabilistic finite-state machine task model was introduced, which when augmented with transition time distributions as well as placement accuracy distributions, results in a compact triad representing the performance of the coupled system-user pair. The applicability of this compact triad towards investigating the deconvolution of user, feedback, learning, and fatigue components of performance was discussed, and future extensions presented. Furthermore, interesting results arose not only from the task-based but also from the questionnaire-based evaluation: for example, an ordering of desirability of a number of application areas for tele-operation arose as a conjecture – showing that most people strongly agree on the application of tele-operation for space or dangerous environments, but that there is potentially strong disagreement for a group of people regarding child instruction through tele-operated robots.

In conclusion, through the presentation of a real-world system and a framework for modeling and evaluation, and through a task- as well as a questionnaire-based evaluation study, valuable results and insights were derived, towards the wider beneficial application of tele-operated robotics by untrained humans in an increasing range of application areas.

Acknowledgment

The authors would like to thank Amr Elfaham, Walid Soulakis, Aziz El Kayoumi, Saif Al Ketbi, and all of UAEU students and IRML friends for their invaluable help, as well as Dr. Eyad Abed for his kind support.

References

[1] J. Cui et al., –A review of teleoperation system control", Proceedings of the 2006 Florida Conference Recent Advances in Robotics (FCRAR), Florida Atlantic University, FL.

[2] D. Kim, J. Kim, K. Lee, C. Park, J. Song, D. Kang. –Excavator tele-operation system using a human arm", Elsevier Automation in Construction, vol. 18, p. 173-182, 2009.

[3] D. Matsui, T. Minato, K. F. MacDorman, H. Ishiguro, –Generating Natural Motion in an Android by Mapping Human Motion", Humanoid Robots, Human-like Machines, 2007.

[4] A. Alissandrakis, N. Otero, J. Saunders, K. Dautenhahn and C. L. Nehaniv, –Helping Robots Imitate — Metrics And Computational Solutions Inspired By Human-Robot Interaction Studies". In J. Gray and S. Nefti-Meziani (Eds.), Advances in Cognitive Systems, IET Press.

[5] A. Alissandrakis, C. L. Nehaniv and K. Dautenhahn, Correspondence Mapping Induced State and Action Metrics for Robotic Imitation. IEEE Transactions on Systems, Man, & Cybernetics, Part B: Cybernetics, Special issue on Robot Learning by Observation, Demonstration and Imitation, Vol. 37, Issue 2, pp. 299-307, 2007

[6] Questionnaire available at Interactive Robots and Media Lab website, http://irml.uaeu.ac.ae (last retrieved January 2010)

[7] Videos of Interactive Robots and Media Lab projects online at: http://www.youtube.com channel irmluaeu

[8] N. Mavridis and D. Hanson, "The IbnSina Center: An Augmented Reality Theater with Intelligent Robotic and Virtual Characters", 18th IEEE Symp. on Human and Robot Interactive Comm. Ro-MAN 2009