

A Methodology to Assess the Acceptability of Human-Robot Collaboration Using Virtual Reality

Vincent Weistroffer, Alexis Paljic, Lucile Callebert, Philippe Fuchs

▶ **To cite this version:**

Vincent Weistroffer, Alexis Paljic, Lucile Callebert, Philippe Fuchs. A Methodology to Assess the Acceptability of Human-Robot Collaboration Using Virtual Reality. 19th ACM Symposium on Virtual Reality Software and Technology, Oct 2013, Singapour, Singapore. pp.39-48, 10.1145/2503713.2503726 . hal-00875775

HAL Id: hal-00875775

https://minesparis-psl.hal.science/hal-00875775

Submitted on 22 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Methodology to Assess the Acceptability of Human-Robot Collaboration Using Virtual Reality

Vincent Weistroffer* Mines ParisTech Robotics Center - CAOR

Alexis Paljic† Mines ParisTech Robotics Center - CAOR

Lucile Callebert‡ Mines ParisTech Robotics Center - CAOR

Philippe Fuchs§ Mines ParisTech Robotics Center - CAOR

©ACM 2013. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in VRST 2013, http://dx.doi.org/10.1145/2503713.2503726.

Abstract

Robots are becoming more and more present in our everyday life: they are already used for domestic tasks, for companionship activities, and soon they will be used to assist humans and collaborate with them in their work. Human-robot collaboration has already been studied in the industry, for ergonomics and efficiency purposes, but more from a safety than from an acceptability point of view. In this work, we focused on how people perceive robots in a collaboration task and we proposed to use virtual reality as a simulation environment to test different parameters, by making users collaborate with virtual robots. A simple use case was implemented to compare different robot appearances and different robot movements. Questionnaires and physiological measures were used to assess the acceptability level of each condition with a user study. The results showed that the perception of robot movements depended on robot appearance and that a more anthropomorphic robot, both in its appearance and movements, was not necessarily better accepted by the users in a collaboration task. Finally, this preliminary use case was also the opportunity to guarantee the relevance of using such a methodology — based on virtual reality, questionnaires and physiological measures — for future studies.

CR Categories: I.2.9 [Artificial Intelligence]: Robotics—Manipulators H.5.2 [Information Interfaces and Presentation]: User Interfaces—Ergonomics, Evaluation/methodology;

Keywords: human-robot collaboration, acceptability, human factors, virtual reality

1 Introduction

Robots are becoming more and more present in our everyday life. Some robots can already handle domestic tasks autonomously, others are used as companions for the youth or the elderly, or even as guides in museums. In a near future, robots will be able to handle more complicated tasks, to assist humans in their everyday work and to collaborate with them.

Human-robot collaboration has already been studied for many years in the industry: specific robots are used to assist workers on difficult tasks, for example on assembly lines. The assets of this collaboration are improved ergonomics for the workers and improved competitiveness for the industry. For now, the main issue being studied has been safety: since physical interaction between robots and workers is needed, the safety of people around robots has to be guaranteed. This issue has been dealt with thanks to the development of new sensors placed directly on robots or in their environments, in order to detect intrusion in a safety perimeter and avoid collision if necessary.

However, another issue, less studied but nonetheless important, is the acceptability of such a collaboration between humans and robots. From the users' point of view, the introduction of robots in their everyday work may be perceived and accepted differently than expected. This acceptability is difficult to evaluate and may be influenced by several factors that are hard to determine.

This is where virtual reality may provide interesting insights and functionalities to study the acceptability of human-robot collaboration. By making users collaborate with virtual robots under different conditions, it should be possible to evaluate the influence of different factors on this acceptability, in a more practical way than if the experiments were performed with real robots.

In our work, we propose to use virtual reality as a simulation environment to study different scenarios of human-robot collaboration and to evaluate their acceptability from the users' point of view. We provide a methodology to assess the acceptability of human-robot collaboration systems, based on questionnaires and physiological measures. This methodology was applied to a preliminary use case to evaluate the impact of robot appearance and movements on the users' perception. A user study was performed to assess the relevance of our methodology in this particular use case.

In Section 2, we describe related work on the acceptability of human-robot collaboration, the methodologies used to evaluate it and the need to use virtual reality in this context. In Section 3, we present the tools we developed to set up our methodology, dedicated to the simulation environment and the acceptability criteria. In Section 4, we describe the design of the use case we implemented to study human-robot collaboration and we explain the user study we performed. Finally, in Section 5, we present the results and the analysis of this user study before concluding on the relevance of our methodology and our future work in Section 6.

2 Related Work

Human-robot collaboration has been studied for a long time, both in the fields of service robotics, for example with the Care-O-bot [Parlitz et al. 2008], and industrial robotics, with European projects such as PiSA [Krüger et al. 2010] and SMErobot [Nillson et al. 2005]. Studies on human-robot collaboration have often focused on safety: since physical interaction between humans and robots is needed, the safety of this interaction has to be guaranteed. In that context, several improvements have been made in safe obstacle avoidance [Pedrocchi et al. 2009], in the safety evaluation of robots [Haddadin et al. 2007] or in intrusion detection [Lenz 2011].

*e-mail: vincent.weistroffer@mines-paristech.fr
†e-mail: alexis.paljic@mines-paristech.fr
‡e-mail: lucile.callebert@mines-paristech.fr
§e-mail: philippe.fuchs@mines-paristech.fr

If the safety of collaborative robots is nowadays well studied, another issue is being less considered: the acceptability of such a collaboration between humans and robots. The notion of acceptability is not easy to define. Nielsen [Nielsen 1993] proposed a model in which acceptability is divided between social acceptability and practical acceptability, the latter gathering the notions of utility and usability (see Figure 1). We tried to apply this abstract definition to our specific context: human-robot collaboration. Social acceptability concerns how the society perceives robots in everyday life, while practical acceptability relates to how people perceive robots when interacting directly with them. In our work, we were especially interested in the practical acceptability of robots from the users' point of view: we tried to study the usability and the satisfaction of people when collaborating with robots.

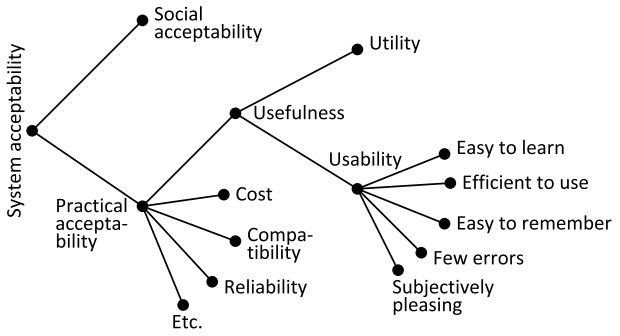


Figure 1: *Nielsen's model of system acceptability.*

The study of the acceptability of human-robot collaboration focuses on several parameters: robot appearance [Goetz et al. 2003], robot movements [Nonaka et al. 2004; Kulić and Croft 2007; Huber et al. 2008; Kupferberg et al. 2011], robot behavior, robot speed, robot safety distance [Karwowski and Rahimi 1991]... When dealing with only one kind of parameter or with a specific system, the implementation of experiments in a real environment with a real robot is the usual way to proceed and may be sufficient. However, if the experiments become more complicated (using several robots for comparison, testing several parameters), the implementation becomes impossible for practical and cost reasons. That is why the use of virtual reality is interesting in that context: it can ease the implementation of the study and enable a large number of test parameters.

In the literature, very few people made use of virtual reality tools to study human-robot collaboration. Most studies used real scenarios where the users only observed robots and did not interact with them. Those were scenarios in which the robots performed specific movements or tasks in front of the users, such as moving from point to point [Shibata and Inooka 1998] or manipulating objects close to the users [Kulić and Croft 2007]. Some studies did involve the users in an interaction task with real robots [Huber et al. 2008], but they were often performed to assess the efficiency of the system, and not the acceptability. The only experiments that used virtual reality to study human-robot collaboration took little advantage of its interaction functionalities. Nonaka et al. [Nonaka et al. 2004] and Inoue et al. [Inoue et al. 2005] used virtual robots (with head-mounted displays or CAVE systems) to evaluate the effect of robot movements and robot proximity on the users, but did not study any interaction and collaboration between them. On a different topic, De Santis et al. [De Santis et al. 2008] used virtual reality to evaluate the usability of different (virtual) robotic arms mounted on a wheelchair, but this was especially for efficiency purposes. Our aim is to benefit from all the functionalities of virtual reality (immersion and interaction), by implementing interactive scenarios in which users collaborate with virtual robots, and by evaluating the acceptability of that collaboration.

The main issue when dealing with acceptability is how to evaluate it. This question is not easy since acceptability is a subjective notion that people often seek to measure in an objective way. The most common tool to evaluate acceptability is to use questionnaires. Different kinds of questionnaires exist in the literature: some of them focus on the robot's attributes [Shibata and Inooka 1998] (human-like, pleasant, interesting...), while others focus on the user's emotions [Nonaka et al. 2004] (surprise, fear...). Bartneck et al. [Bartneck et al. 2009] performed a complete review on questionnaires about human-robot interaction and proposed their own questionnaires intended to be used by the community of robot developers. Even if questionnaires can be analysed directly, they rely on subjective impressions of the users and their results should be correlated with more objective measures. One solution is user observation during the interaction: Kanda et al. [Kanda et al. 2004] defined an evaluation score based on body movements, Kooijmans et al. [Kooijmans et al. 2007] implemented a software to record and analyse data from video clips, Minato et al. [Minato et al. 2005] tried to analyse eye gaze during the interaction. Those studies showed that defining an objective way to observe and evaluate the user's behavior is definitely not an easy method. Another solution is to evaluate the user's affective state thanks to physiological measures. The main idea in the literature [Rani et al. 2002; Sarkar 2002; Kulić and Croft 2007] was to use multiple physiological signals (such as heart rate, skin conductance, facial muscle contraction) in order to infer human physiological states (stress, anxiety) and enable the robot to react appropriately. The same idea could be used to assess the user's acceptability towards human-robot collaboration.

As a summary, the scope of our work is to study the acceptability of human-robot collaboration from the users' point of view. Our approach is to use virtual reality to make users collaborate with virtual robots and gather their impressions for different conditions. We propose to use a combination of questionnaires and physiological measures to evaluate the acceptability of the collaboration in an objective way.

3 Methodology Tools

In this section, we describe the tools that we developed and used to set up our methodology. Since our approach was to use virtual environments to study human-robot collaboration, a large part of this section is dedicated to explaining how to model and control virtual robots. We also describe the criteria we chose to evaluate the acceptability of human-robot collaboration: questionnaires for subjective impressions and physiological measures for estimating human affective state.

3.1 Simulation Environment

3.1.1 Robot Models

In our work, we focused on robotic articulated arms: their structure is often represented by a series of rigid bodies linked to each other by joints. We considered only rotational joints: the state of a joint is represented by an angle around a specific axis. The adaptation of this representation to virtual robots is straightforward: the virtual robot is modelled as a hierarchy of 3D frames (the joints of the robot) and 3D objects as their specific children (the rigid bodies). Thanks to this hierarchy, rotating a specific joint moves the entire arm from this joint to the end-effector. Examples of virtual robotic arms are presented in Figure 2.

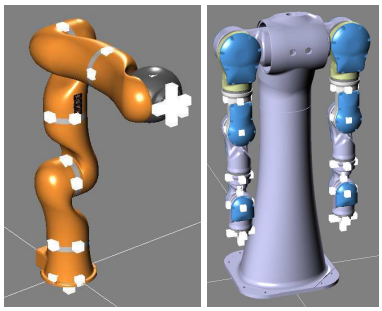


Figure 2: *Examples of virtual robotic arms. The 3D frames are visible in white color.*

3.1.2 Robot Kinematics

Before collaborating with robots in virtual environments, it is necessary to know how to control them. By modifying directly the joints states of a robotic arm, it is easy to find the cartesian state (position, orientation) of its end-effector(s). This process is called forward kinematics. Even if forward kinematics is the most direct way to control a robotic arm, it is not the most usual and practical way to proceed: end-effectors are often assigned target positions that they have to follow. Therefore, given the cartesian states of the targets, the joints states of the arm have to be computed to give the correct positions to the end-effectors. This process is called inverse kinematics.

Inverse kinematics have been well studied both in the fields of animation (for human arms and skeletons) and robotics. Several methods have been proposed to solve this complex problem, that can be classified between analytical and numerical solutions. Analytical solutions are difficult to produce and depend largely on the structure of the arm [Kucuk and Bingul 2006]. Numerical solutions, based on optimisation or jacobian methods, are easier to implement and can be generalised to different robotic arm structures. Buss and Kim [Buss and Kim 2005] presented a complete review on the methods using jacobians and provided their source code online [1].

As Buss' algorithms are dedicated to control multibodies for computer graphics, with multi-targets and in real time, we decided to use them and to integrate them in our system to control virtual robotic arms. Buss' model uses a tree to represent the structure of a rigid multibody (such as a robotic arm): the nodes of the tree represent the joints, while the leaves are the end-effectors (see Figure 3). Each node of the tree gathers all the needed data about the corresponding joint: joint state (angle of rotation, axis of rotation), cartesian state (absolute position, relative position to the former joint). The states of the nodes are updated in real time according to the targets' positions, thanks to inverse kinematics algorithms based on jacobian methods.

In our system, each of our robot models is coupled with a tree structure in real time (see Figure 4). This tree structure is first built according to the robot's hierarchy (initialisation phase). The robot's hierarchy (series of joints, joints initial positions and axes) is stored in a text file (identification file) that is used as a parameter for the construction of the tree. The nodes data are then updated in real time depending on the targets' positions. The joints states are finally sent to the robot's model for graphical update (mapping phase). With this principle, the core of the inverse kinematics algorithms is done in the tree and the robot's model is only updated according to the nodes data.

[1] `http://www.acm.org/jgt/papers/BussKim05`

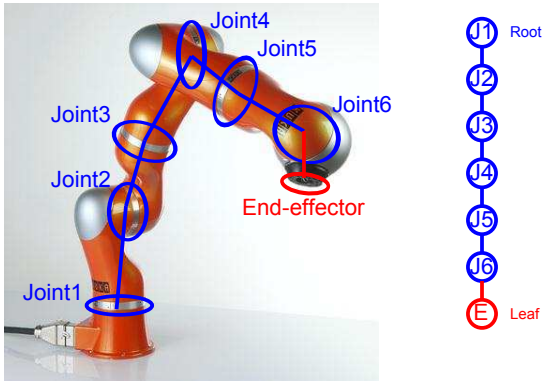


Figure 3: *Example of a tree structure for a robotic arm (represented here with six degrees of freedom).*

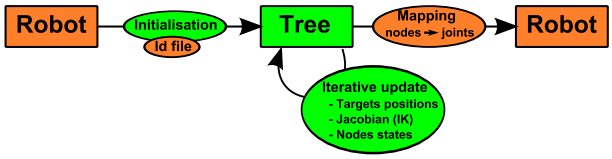


Figure 4: *Principle of the coupling between a robot structure (in orange) and a tree structure (in green).*

We used this architecture with several robots. Each robot had a different identification file, since each robot had a different structure and thus was coupled with a different tree. This system was tested successfully with predefined trajectories, with targets generated and controlled by a mouse cursor, and finally with targets generated by users in a virtual environment: thanks to motion capture, the users were able to guide the end-effectors of robotic arms with their hands, in real time (see Figure 5). With this principle, it was possible to give different behaviors to robots by generating different movements profiles with the targets.

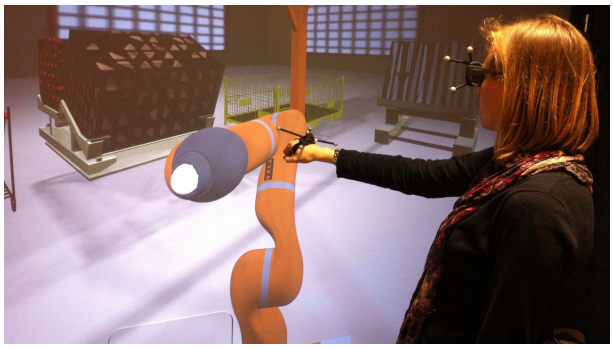


Figure 5: *Robot control in a virtual environment using inverse kinematics and motion capture.*

3.2 Acceptability Criteria

One of the aims of our methodology was to find the relevant criteria for the evaluation of acceptability that enabled to show that a certain situation was better accepted than others. In that context, we chose to use questionnaires, for subjective impressions, and physiological measures, for correlation with the questionnaires through the estimation of human affective state.

3.2.1 Questionnaires

As a starting point for our questionnaires, we focused on the 'Godspeed questionnaires' proposed by Bartneck et al. [Bartneck et al. 2009]. Those are five consistent questionnaires using semantic differential scales, on the topics of anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots, based on empirical studies and dedicated to robot developers. Those questionnaires were adapted to the use cases we wanted to study and were mainly focused on anthropomorphism, perceived intelligence and perceived safety.

3.2.2 Physiological Measures

In the literature, studies with physiological measures for human-robot interaction usually used heart rate and skin conductance. Kulić and Croft [Kulić and Croft 2007] also used facial muscle contraction but they concluded that it was not a good indicator for estimating human affective state.

Skin conductance level (SCL) measures the skin sweating activity of the user and is directly related to his stress level: a higher SCL usually means a higher level of stress. Since SCL differs between users, it is necessary to always compare it to an average or control value (specific to each user). Skin conductance responses (SCR) may also occur: they are peaks in SCL that happen spontaneously one to three times per minute (depending on users) or directly after a stimulus [Dawson et al. 2000]. The amplitude of SCR may be related to the strength of the reaction to the stimulus. Heart rate activity is usually more difficult to analyse, since it can be influenced by several factors (physical state, fatigue, body position).

For our studies, the average SCL, the frequency of SCR and the average heart rate (HR) seemed to be the relevant criteria to estimate human affective state and to evaluate acceptability in a collaboration task. It was also important to use physiological signals that were simple to handle, whose measurements were non-invasive and allowed the users to move and interact in a virtual environment, which is the case with HR and SCL.

4 Preliminary Study

We applied the methodology described in Section 3 to a preliminary use case. The aim of this study was to evaluate the effect of robot appearance and movements on the acceptability of a simple collaboration task.

4.1 Use Case Design

4.1.1 Collaboration Task

A simple use case was implemented for collaboration between a user and a robot. We wanted this use case to be interactive, to give specific roles to both the robot and the user, but also to be simple enough to allow the user to look at the robot's appearance and movements (and not to focus on the task only). It was also important that this use case could be related to a real industrial task.

In our study, we took inspiration from an industrial task in which a robot may have to hand mechanical parts to a worker on an assembly line. For abstraction purposes, this task was simplified into a cubes exchange process between a user and a robot (see Figure 6). The user and the robot were on each side of a table. Eight cubes (four red and four blue) were positioned in front of the robot (unreachable by the user), while storage areas were placed in front of the user (unreachable by the robot). The robot had to take cubes one by one and give them to the user, who had to store them in the correct area, depending on the color of the cubes (blue on the left or red on the right).

The robot was controlled thanks to the inverse kinematics process exposed in Section 3.1.2: the end-effector of the robot followed a specific target in the virtual environment. The movements of the target were trajectories between a neutral position P_n, the positions of the cubes $P_{c1}...P_{c8}$ and an interaction position P_i from which the cubes were exchanged with the user (see Figure 6). The movements were divided in two steps: $P_n \rightarrow P_c \rightarrow P_i$ (taking a cube and waiting for the user to grab it), and $P_i \rightarrow P_n$ (coming back to the neutral position).

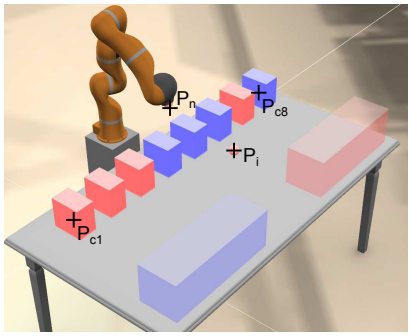


Figure 6: *The collaboration task and the robot collaboration steps.*

To give some challenge to the collaboration task, the placement of the cubes on the table was randomised, so as the order in which they were taken by the robot. The task was considered done when all the cubes were put in the storage areas.

4.1.2 Robots Conditions

Different robots were used to perform the collaboration task (see Figure 7). We chose industrial robots whose appearances differed in anthropomorphism and safety clues. Robot R0 (a Universal UR10) is a simple, light and flexible robot. Robot R1 (a Kuka KR5SI) is an industrial robot surrounded with safety pads for intrusion/collision detection. Robot R2 (a Kuka LWR4) is specifically dedicated to interaction with humans and reproduces the human arm structure. Finally, robot R3 (a Motoman SDA10) is a humanoid robot with two arms.

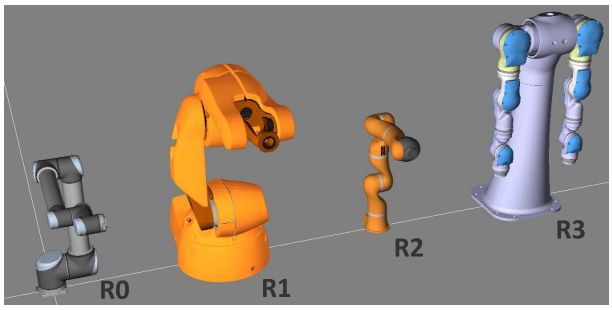


Figure 7: *The robot models used for the collaboration task.*

4.1.3 Movements Conditions

We wanted each robot to move with different movements profiles: machinelike [M1] and humanlike [M2]. Usually, humanlike movements are expected to have different trajectories, different speed

profiles, but also more variability when repeating actions and reaching positions. It seemed interesting to study the effect of humanlike movements on the acceptability of the collaboration task.

Since each robot was controlled thanks to inverse kinematics (IK) algorithms, M1 and M2 movements were generated by giving different profiles to the IK target's movements. The M1 movements were straight lines at constant speed between the specified positions. To generate the M2 movements, we recorded the hand trajectories and speed of a real user performing the specific movements in a virtual environment: the user was placed on the robot side of the table and performed a pointing task between the specified positions. For this recording phase, P_n and P_i were represented as virtual spheres in the environment: the sphere for P_n was 5cm wide, while the sphere for P_i was 20cm wide to give the movements some variability in the reaching position. A complete set of 16 movements was recorded (2 steps for each cube) and could be replayed at any moment in the virtual environment (see Section 4.1.4 for more details on the implementation). These movements were applied to the inverse kinematics targets so that the robots could follow them in real time. The speed of M1 movements were adapted to match the average speed of M2 movements.

Since robot R3 had two arms, slightly different movements were generated for this model. The same collaboration steps than for the other robots were used, except that the cubes on the right side of the table were taken by the right arm, while the cubes on the left side were taken by the left arm. In order to prevent the arms from colliding with each other, a slight offset was given to the positions of P_n and P_i: 5cm on the right and on the left, for each arm respectively. A different set of M2 movements was also recorded for this robot, taking the offset positions of P_n and P_i into account.

The main challenge when recording the M2 movements was to reduce the transition effects between each step: since each movement was recorded separately, the transitions between them could occur in a non-smooth manner during playback. This was overcome by always coming back to the same neutral position P_n and by recording the same step several times and choosing the best.

4.1.4 Use Case Implementation

This use case was implemented thanks to 3DVIA Virtools software[2]. The virtual environment was rendered on a back-projected wall (3.1m $\times$ 1.7m), with active stereoscopy. We used infrared ART[3] cameras for motion capture: both the head and the dominant hand of the user were tracked in real time thanks to passive targets. The motion capture data were sent to the Virtools software through VRPN[4]. Specific implementations were performed inside the Virtools software: virtual robots' control through inverse kinematics, movements recording and movements playback (the VRPN stream was recorded/replayed thanks to VRPN logging and playback functionalities). Figure 8 shows the entire settings of this use case.

4.2 User Study Design

The aim of our work was to study the influence of robot appearance and robot movements on the acceptability of the collaboration task, from the users' point of view. First, we wanted to know what was the impact of robot appearance between R1, R2 and R3. Secondly, we wanted to know if the users perceived a difference between M1 and M2, if this difference was the same for each robot and in which way it influenced acceptability.

[2] http://www.3ds.com/products/3dvia/3dvia-virtools/
[3] http://www.ar-tracking.com/home/
[4] http://www.cs.unc.edu/Research/vrpn/

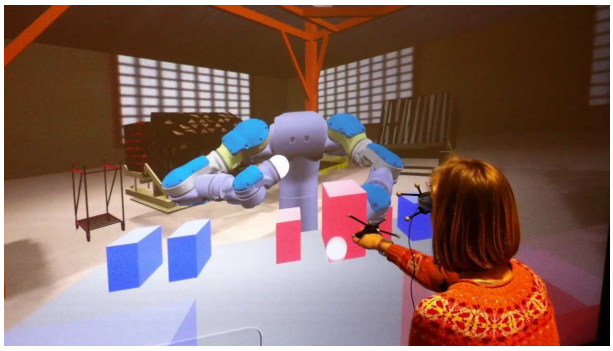


Figure 8: *The implementation of the use case in a virtual environment.*

In order to compare each condition, we performed a user study on the collaboration task. The independent variables were robot appearance (R1, R2, R3) and robot movements (M1, M2). A within-subject design was chosen: each subject tested the 6 conditions. For each condition, the collaboration task was repeated three times.

The different steps of our user study were the following (see Table 1 for an example). For each subject, the test began with a training phase with robot R0 and movement M1 to let the user familiarise with the collaboration task. The test was then divided in three steps corresponding to each robot condition: for each step, the user was shown a specific robot and had to complete the collaboration task with both movements conditions. The order of robots conditions was randomised, and the order of movements conditions was randomised for each robot as well. At the end of each step, the user was asked to fill out a questionnaire [Q_M] to assess the perceived differences between both movements. Finally, at the end of the whole test, the user was asked to fill out an additional questionnaire [Q_R] to assess the perceived differences between the robots' appearances.

This design was chosen to enable the users to compare directly the movements conditions for a specific robot: since the differences between M1 and M2 could be subtle, it was better to group those conditions next to each other and to ask the Q_M questions directly after the robot condition, in order to help user memory. The Q_R questions were asked at the end of the test once the user had seen each robot.

4.3 User Study Measures

4.3.1 Questionnaires

Two questionnaires were used: one to compare the two movements types for a given robot (Q_M, see Table 2), one to compare the robots' appearances between them (Q_R, see Table 3). For both questionnaires, we used five questions for which the users had to rank each condition on a 6-point Likert scale. The questions dealt with anthropomorphism, competence, relaxation, safety and impression.

In order to simplify the questionnaire process during the user study, the questions were displayed on the screen (one by one) and the user had to give his answers verbally to the coordinators.

4.3.2 Physiological Measures

To strengthen the questionnaires results, our methodology focused on estimating human affective state by measuring heart rate and skin conductance. We used a physiological monitoring system from

Step	Training	Step 1			Step 2			Step 3			
Condition Repetitions	R0M1 1	R3M1 3	R3M2 3	Q_M	R1M2 3	R1M1 3	Q_M	R2M2 3	R2M1 3	Q_M	Q_R

Table 1: *An example of the different steps of the user study.*

Questions	Propositions			Answers M1	Answers M2
The movements seemed rather...	Machinelike	1 - 6	Humanlike		
The movements seemed rather...	Incompetent	1 - 6	Competent		
Did you feel rather...	Anxious	1 - 6	Relaxed		
Did you feel rather...	Unsafe	1 - 6	Safe		
What is your global impression	Dislike	1 - 6	Like		

Table 2: *The Q_M questionnaire to grade and compare the different movements for a specific robot.*

Biopac Systems[5] with a Bionomadix PPG (Photoplethysmogram) and EDA (Electrodermal Activity) amplifier. PPG signal provides information on the subject's heart rate, while EDA signal measures the skin conductance level (SCL). Those signals were measured on the fingertips of the non-dominant hand of the users (keeping the dominant hand available for interaction). The system was wireless: a small transmitter was placed on the user's forearm and sent the data to a receiver (connected to a computer through ethernet) for storage and later analysis.

PPG and EDA signals were sampled at 250Hz, and were recorded and analysed thanks to AcqKnowledge software (from Biopac Systems). Skin conductance responses (SCR) were detected thanks to this software: they were peaks in SCL with a maximum duration of 5s and a minimum amplitude of 0.02μS. Events (condition start, task start and task stop) were marked by the experimentators in real time, with a key press, and stored with signal curves. Those events markers simplified the offline segmentation of the signals.

For each condition, the physiological criteria we chose to measure were: the average heart rate (HR), the average skin conductance level (SCL), the frequency of skin conductance responses (SCR). Those criteria were computed for each task of the test. To allow comparison between all the users, those criteria were normalised (between 0 and 1) according to the subject's corresponding minimum and maximum values over all the tasks of the test. For a specific condition, the average over the three corresponding tasks was taken as the final value for each criterion.

5 User Study Results

A total of 20 subjects participated in the user study. The average age of the participants was 26. For each subject, the test duration was approximately 45 minutes, including the questionnaires. We first present the results and analysis of the questionnaires. Then, we describe the correlation of those results with the physiological measures.

5.1 Questionnaires

5.1.1 Robot Appearance

The Q_R questions on robot appearance were asked at the end of the test, once the subject had completed the collaboration task with each robot. The robots' appearances were scored on 6-point Likert scales, on the following criteria: anthropomorphism, competence, relaxation, safety and impression. Kruskal-Wallis (KW) tests were performed to evaluate if robot appearance was a significant factor for each criterion, and single Mann-Whitney (MW) tests were used for condition pairs analysis. The values mentioned in the following results are the average over all the users.

The main difference to be seen between the robots' appearances was the anthropomorphism dimension (see Figure 9). With its two arms, robot R3 was perceived as the most humanlike (4.8). Robot R2 was more perceived like an animal (3.8), while robot R1 was definitely seen as an industrial robot (2). A KW test confirmed that robot appearance was indeed a significant factor for anthropomorphism ($KW(2) = 29.7, p < 10^{-6}$) and MW tests showed that the difference was significant for each condition pair.

For the relaxation, safety and impression dimensions, robot R1 always got the lowest answers (less than 3.5), while robots R2 and R3 got similar results (more than 4.5 for each criterion). A KW test confirmed a significant influence of robot appearance on those criteria ($KW(2) = 9.4, p = 0.01$; $KW(2) = 11.6, p = 0.003$; $KW(2) = 16.7, p < 10^{-3}$; respectively). MW tests showed this influence was mainly due to the low scores of robot R1 but that there was no big difference between robots R2 and R3. However, it can be noted that robot R2 got slightly better results than robot R3 on those criteria (see Figure 9).

For the competence dimension, each robot was scored globally the same (between 4.1 and 4.7). No significant difference was shown by a KW test ($KW(2) = 4, p = 0.14$).

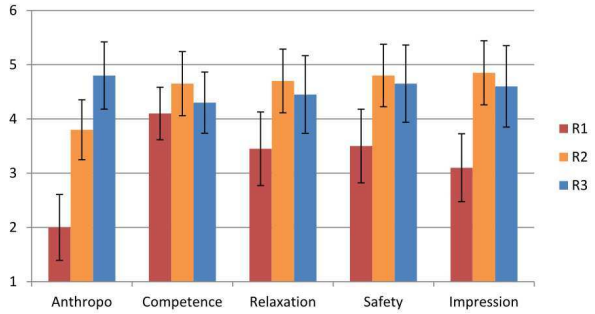


Figure 9: *The average scores of each robot for the appearance questionnaire.*

On the whole, the results of the Q_R questionnaire showed that robot R1 was the least appreciated by the users. According to the users' comments, its tough structure did not inspire trust or safety compared to the other robots. Robots R2 and R3 were globally rated similarly. It is to be noted that even if robot R3 was perceived as

[5] `http://www.biopac.com/`

Questions	Propositions			Answers R1	R2	R3
The robot seemed rather...	Machinelike	1 - 6	Humanlike			
The robot seemed rather...	Incompetent	1 - 6	Competent			
Did you feel rather...	Anxious	1 - 6	Relaxed			
Did you feel rather...	Unsafe	1 - 6	Safe			
What is your global impression	Dislike	1 - 6	Like			

Table 3: *The Q_R questionnaire to grade and compare the different robots' appearances.*

the most humanlike (with its two arms), this was not necessarily a good point for the other criteria: working with a humanoid robot on the collaboration task was often perturbing, and the relevance to use two arms for this task was often questionned. In matters of relaxation, safety or global impression, robot R2 was often preferred to robot R3 because it was better adapted to the task. Therefore, increased anthropomorphism does not necessarily imply a better acceptability.

5.1.2 Robot Movements

The Q_M questions on robot movements were asked once the users had completed the collaboration task with both movements conditions for a specific robot. The users were asked to focus on the robot's movements and to answer the questionnaire according to the difference(s) they might have perceived. For analysis and comparison between each robot, we subtracted the scores of M1 movements (straight uniform) to the scores of M2 movements (human motion capture). For each criterion (anthropomorphism, competence, relaxation, safety and impression), each robot was therefore given a score Δ (negative or positive) to assess the difference in movements perception. As for the questionnaire on robot appearance, KW and MW tests were performed to analyse the results.

First, on the average of all the users, the Δ scores for anthropomorphism were positive for each robot, meaning that M2 movements were consistently perceived as more humanlike than M1 movements. However, the amplitude of Δ was not the same for each robot (1.4 for R1, 0.45 for R2 and 0.2 for R3): the difference was perceived more strongly for robot R1 than for the others (see Figure 10). The physical appearance of robot R1 probably made the difference between M1 and M2 more obvious than for the other robots, whose appearance was more anthropomorphic. A KW test showed indeed a significant influence of robot appearance on movements perception for anthropomorphism ($KW(2) = 4.5, p = 0.1$).

A second interesting point was about the global impression dimension, for which a significant impact was found ($KW(2) = 5.9, p = 0.05$). On the whole, the more the robot was anthropomorphic, the less the M2 movements were appreciated (see Figure 10). Actually, M2 movements were preferred to M1 movements only in the case of robot R1 ($\Delta = 0.35$), while robots R2 and R3 were preferred with M1 movements ($\Delta = -0.1$ and -0.75, respectively). Since robot R1 was perceived as an industrial robot, adding humanlike movements to its behavior was appreciated, whereas it was perturbing with robots R2 and R3, whose appearance was already perceived as humanlike.

No significant difference was shown for the other criteria. However, it can still be noted that, on the average of all the users, M1 movements were perceived as more safe and more competent than M2 movements (the Δ scores for safety and competence were negative; see Figure 10). This means that, in terms of safety and competence, more mechanical movements were better accepted than more natural ones.

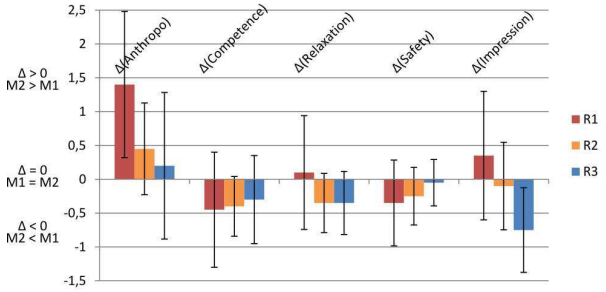


Figure 10: *The average difference between M1 and M2 for each criterion, for each robot.*

5.1.3 Global Analysis of the Questionnaires

The results of the questionnaires showed three interesting points.

The first point is on robot appearance. A difference in anthropomorphism was indeed perceived between each robot, robot R1 being the most industrial and robot R3 the most humanlike. However, robots R2 and R3 were appreciated globally on the same level, meaning that a more anthropomorphic robot is not necessarily better accepted. This may be explained by two reasons: first, robot appearance has to match its function (an industrial task in our case), as shown by Goetz et al. [Goetz et al. 2003]; secondly, a too anthropomorphic robot may fall into the uncanny valley [Mori 1970], as it was sometimes the case with robot R3.

The second point is on the perception of robot movements: our study showed that the appreciation of robot movements depended on the robot's appearance. Indeed, an industrial robot (like R1) may tend to highlight the differences between M1 and M2 movements, while more anthropomorphic robots (like R2 and R3) may gather the attention of the users more on their appearance and less on their movements, thus perceiving less difference between M1 and M2 movements.

Finally, the third point is about the relevance to use humanlike movements on robots for an industrial task. Humanlike movements were better accepted in the case of robot R1 but not for more anthropomorphic robots like R2 and R3. This result shows that there could be a trade-off to establish between appearance and humanlike movements. Two main reasons may account for this fact. First, adding humanlike movements to an already anthropomorphic robot reinforced the uncanny valley effect: especially with robot R3, some users were perturbed and did not know how to consider this robot — as a human or as an industrial robot. Secondly, too much anthropomorphism increased the impression that the robots were not efficient enough: for the collaboration task, the users often preferred M1 movements since they could predict the robot's actions and adapt to the robot's behavior. They were more focused and felt more efficient.

The non-relevance to use humanlike movements (at least with robots R2 and R3) for a better acceptability may seem contradictory with other studies which showed the opposite trend [Kupferberg et al. 2011]. However, the main difference here is about the context of the study: in our work, we focused on an interaction scenario where the users had an active role, whereas the other studies used passive scenarios where the users only observed robot movements without interaction. Therefore, the use case context has a strong impact on the results of acceptability.

5.2 Physiological Measures

We analysed the physiological measures in correlation with the results of the questionnaires. On the whole, a total of three criteria, normalised between 0 and 1, were used for each condition: HR, SCL, frequency of SCR. We performed 2-way ANOVAs (analysis of variance) to detect any significant effect of the robots and movements conditions for the different criteria.

First, a slight effect of the movements conditions could be shown for heart rate ($F(17,1) = 2.75, p = 0.11$; 2 subjects were not taken into the analysis for improper heart rate signal). As can be seen in Figure 11, the average heart rate was always lower for M2 movements with each robot condition. This could mean that the subjects focused more on the task when dealing with more mechanical movements (M1): this correlates with the questionnaires results which showed that M1 movements were perceived as more competent than M2 movements. This difference in heart rate activity between M1 and M2 conditions could also mean that the subjects were more relaxed with M2 movements; this is however not correlated by the questionnaires results.

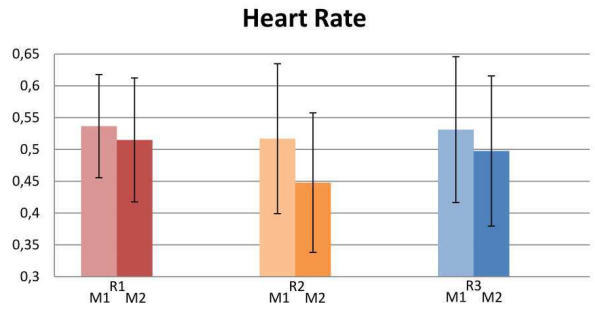


Figure 11: *The average normalised heart rate for each condition.*

A second slight effect was shown between robots conditions for SCL and SCR frequency. Figure 12 shows that SCL was generally higher for robot R1 than for the other robots. This can be explained by the more industrial appearance of robot R1: the subjects were more impressed by this robot than by the others. However, the SCR frequency was higher for robots R2 and R3 than for robot R1 (see Figure 12). This should not be necessarily correlated with a higher level of stress, but rather to a specific reaction to the robots' movements: when moving, robots R2 and R3 induced more responses because of their anthropomorphic appearance. It is also to be noted that the difference in SCR between M1 and M2 for each robot followed the same pattern than the Q_M questionnaire results for impression: there were more SCRs for M2 with robot R1, which was the preferred movements condition for this robot, whereas the opposite trend was seen for robots R2 and R3 (more SCRs for M1 which was the preferred condition).

On the whole, some interesting results were found with physiological measures that correlate with the questionnaires results. We are

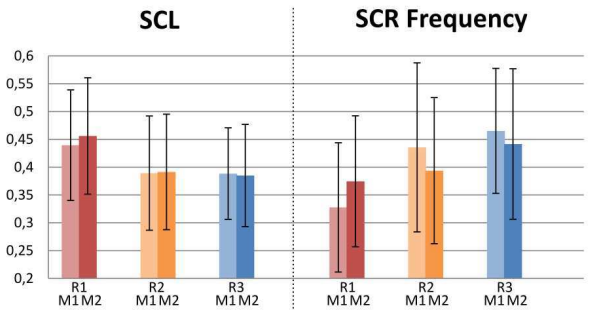


Figure 12: *The average normalised SCL and SCR frequency for each condition.*

conscious that those results are still subtle and that we need more knowledge about how to analyse them correctly. The main challenge of our approach was that we did not use an event-related study (analysing SCR according to specific events in the environment), but rather a study on continuous conditions. However, those first results showed that this approach seems promising.

6 Conclusion and Future Work

The topic of our work was human-robot collaboration. We pointed out that, in the literature, the main focus was put on the physical safety of users in the environment of robots, but that less importance was given to the acceptability of collaboration between humans and robots. We were especially interested to study how people reacted to robots during a collaboration task. The need to use virtual reality in this context was relevant since several parameters had to be tested and such a study would require heavy setups with real robots and robot-specific developments. Our aim was to make users collaborate with virtual robots under different conditions and to gather their impressions.

The first step of our work was to develop and use suitable tools to set up our methodology. We first focused on the simulation environment, with the development of inverse kinematics algorithms, inspired from the literature, which enabled us to control different kinds of virtual robotic arms in a practical way. We also chose the relevant criteria to evaluate acceptability in an objective way for our use cases: using questionnaires and estimating human affective state through heart rate and skin conductance measures.

A simple use case was implemented in which users had to collaborate with virtual robots. The conditions we wanted to compare were robot appearance (three different robots) and robot movements (mechanical movements and movements generated from motion capture). A user study was performed to gather the users' impressions and to evaluate the acceptability of each condition.

The results of this preliminary use case showed that movements perception depended on robot appearance: an industrial robot tended to highlight the differences between mechanical and humanlike movements, whereas anthropomorphic robots gathered the attention on their own appearance. Moreover, a too anthropomorphic robot was not necessarily better accepted: humanlike movements were preferred in the case of industrial robots but not in the case of an anthropomorphic robot. This implies that a trade-off probably has to be established between appearance and humanlike movements. This trend could be due to two main reasons. First, the uncanny valley effect tended to appear when already anthropomorphic robots

moved with natural movements. Secondly, too much anthropomorphism increased the impression that the robots were not efficient enough in the collaboration task. Those results were obtained in an industrial context; they would probably be different in another context, such as in service robotics.

This preliminary study was an attempt to assess the acceptability of human-robot collaboration with a simple use case. We used virtual reality as a simulation environment to implement a collaboration task and study different parameters, and questionnaires and physiological measures to estimate the users' impressions and acceptability of human-robot collaboration. We showed that this methodology was relevant to study human-robot collaboration, even if we still lack some knowledge about the use and analysis of physiological measures. In the future, we intend to apply this methodology to more tangible scenarios, directly inspired from assembly lines in the industry, both in virtual and real environments. Those scenarios will be based on more advanced collaboration and will aim at studying alternative parameters, such as robot adaptive speed and robot safety distance, for which the use of physiological measures may be better suited. We believe that the methodology we proposed will be able to provide relevant results on those use cases.

7 Acknowledgments

The authors would like to thank the subjects that participated in the user study. This research benefited from the support of the Chair 'PSA Peugeot Citroën - Robotics and Virtual Reality' led by MINES ParisTech, and supported by PEUGEOT S.A.. The partners of the Chair cannot be held accountable for the content of this paper, which engages the authors' responsibility only.

References

BARTNECK, C., CROFT, E., AND KULIĆ, D. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics 1*, 1, 71–81.

BUSS, S. R., AND KIM, J.-S. 2005. Selectively damped least squares for inverse kinematics. *Journal of Graphics Tools 10*, 37–49.

DAWSON, M. E., SCHELL, A. M., AND FILION, D. L. 2000. *Handbook of Psychophysiology*. Cambridge University Press, ch. 8, 200–223.

DE SANTIS, A., DI GIRONIMO, G., MARZANO, A., SICILIANO, B., AND TARALLO, A. 2008. A virtual-reality-based evaluation environment for wheelchair-mounted manipulators. *Eurographics Italian Chapter Conference 1*, 1–8.

GOETZ, J., KIESLER, S., AND POWERS, A. 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*, 55–60.

HADDADIN, S., ALBU-SCHÄFFER, A., AND HIRZINGER, G. 2007. Safety evaluation of physical human-robot interaction via crash-testing. In *In Robotics: Science and Systems Conf. (RSS2007)*, 217–224.

HUBER, M., RICKERT, M., KNOLL, A., BRANDT, T., AND GLASAUER, S. 2008. Human-robot interaction in handing-over tasks. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, 107–112.

INOUE, K., NONAKA, S., UJIIE, Y., TAKUBO, T., AND ARAI, T. 2005. Comparison of human psychology for real and virtual mobile manipulators. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, 73–78.

KANDA, T., ISHIGURO, H., IMAI, M., AND ONO, T. 2004. Development and evaluation of interactive humanoid robots. *Proceedings of the IEEE 92*, 11 (nov.), 1839–1850.

KARWOWSKI, W., AND RAHIMI, M. 1991. Worker selection of safe speed and idle condition in simulated monitoring of two industrial robots. *Ergonomics 34*, 5, 531–546.

KOOIJMANS, T., KANDA, T., BARTNECK, C., ISHIGURO, H., AND HAGITA, N. 2007. Accelerating robot development through integral analysis of human-robot interaction. *Robotics, IEEE Transactions on 23*, 5 (oct.), 1001–1012.

KRÜGER, J., KATSCHINSKI, V., SURDILOVIC, D., AND SCHRECK, G. 2010. Flexible assembly systems through workplace-sharing and time-sharing human-machine cooperation (PISA). *Robotics (ISR), 2010 41st International Symposium on and 2010 6th German Conference on Robotics (ROBOTIK) 1* (june), 1–5.

KUCUK, S., AND BINGUL, Z. 2006. *Robot Kinematics: Forward and Inverse Kinematics*. Industrial Robotics: Theory, Modelling, Control, ch. 4, 117–148.

KULIĆ, D., AND CROFT, E. 2007. Affective state estimation for human-robot interaction. *Robotics, IEEE Transactions on 23*, 5 (oct.), 991–1000.

KUPFERBERG, A., GLASAUER, S., HUBER, M., RICKERT, M., KNOLL, A., AND BRANDT, T. 2011. Biological movement increases acceptance of humanoid robots as human partners in motor interaction. *AI & Society 26*, 339–345.

LENZ, C. 2011. *Context-aware human-robot collaboration as a basis for future cognitive factories*. PhD thesis, Technische Universität München.

MINATO, T., SHIMADA, M., ITAKURA, S., LEE, K., AND ISHIGURO, H. 2005. Does gaze reveal the human likeness of an android? In *Development and Learning, 2005. Proceedings. The 4th International Conference on*, 106–111.

MORI, M. 1970. The uncanny valley. *Energy 7*, 33–35.

NIELSEN, J. 1993. *Usability engineering*. Academic Press, Boston.

NILLSON, K., JOHANSSON, R., ROBERTSSON, A., BISCHOFF, R., BROGÅRDH, T., AND HÄGELE, M. 2005. Productive robots and the SMErobot project. In *Book of Abstracts of Third Swedish Workshop on Autonomous Robotics*.

NONAKA, S., INOUE, K., ARAI, T., AND MAE, Y. 2004. Evaluation of human sense of security for coexisting robots using virtual reality. 1st report: evaluation of pick and place motion of humanoid robots. In *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, vol. 3, 2770–2775.

PARLITZ, C., HÄGELE, M., KLEIN, P., SEIFERT, J., AND DAUTENHAHN, K. 2008. Care-o-bot 3 - rationale for human-robot interaction design. In *In Proceedings of 39th International Symposium on Robotics (ISR), Seul, Korea*.

PEDROCCHI, N., MALOSIO, M., AND TOSATTI, L. 2009. Safe obstacle avoidance for industrial robot working without fences.

In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 3435–3440.

RANI, P., SIMS, J., BRACKIN, R., AND SARKAR, N. 2002. Online stress detection using psychophysiological signals for implicit human-robot cooperation. *Robotica 20*, 673–685.

SARKAR, N. 2002. Psychophysiological control architecture for human-robot coordination - concepts and initial experiments. In *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, vol. 4, 3719–3724.

SHIBATA, S., AND INOOKA, H. 1998. Psychological evaluations of robot motions. *International Journal of Industrial Ergonomics 21*, 6, 483–494.