A NOVEL HUMAN-MACHINE INTERACTION ARCHITECTURE – INTENTION RECOGNITION APPROACH

Karim A. Tahboub

College of Engineering and Technology
Palestine Polytechnic University
P. O. Box 198
Hebron - Palestine

Abstract: A novel human-machine interaction architecture is presented. It is based on the machine intention recognition of the human. This work is motivated by the desire to minimize the need for classical direct human-machine interface and communication. Here, the intention-action-state scenario is modified and modelled by dynamic Bayesian networks to facilitate for probabilistic intention inference. The recognized intention, then, drives the interactive behaviour of the machine such that it complies with the human intention in light of the real state of the world. An illustrative example of a human commanding a mobile robot remotely is given and discussed in details. *Copyright © 2005 IFAC*

Keywords: Intention recognition; human-machine interaction; human-robot interaction

1. INTRODUCTION

In many multi-agent systems, when two agents or more cooperate or compete to achieve a certain task, the problem of recognizing the intention of others arises. Usually, the lack to a reliable and full communication of intentions calls for inferring the intention from the actions or even from changes in the environment due to agent actions.

Consider the example of a human operator sitting on and commanding a robotic wheelchair with extended perceptual and actuating capabilities. The robotic wheelchair can navigate autonomously in the environment. However, because of the interaction with the human, a semi-autonomous system results; the wheelchair senses the environment and tries to comply with the human commands while protecting her from any collision with external bodies. In other words, the wheelchair gives up a part of its autonomy in favour of better cooperation with the human user. However, without recognizing the real intentions of the human, the wheelchair can behave in a less-cooperating and a more-frustrating way. On the other hand, it is impractical and annoying to ask the human operator to communicate her intentions explicitly either by speech or through other interfacing devices. Thus, it is favourable that the wheelchair itself infers the actions (commands or their effects) of the human into intentions (Tahboub, 2001). There are many other applications that call for intelligent intention recognition to improve the human-machine interaction including intelligent cars and intelligent transportation systems, physical training machines, disabled-people assistive technologies, and home appliances.

In the literature, intention recognition and plan recognition are used interchangeably while meaning the same thing. However, in this work, a distinction is made between intentions and plans. Whilst philosophers like Bratman (Bratman, 1999) consider intention as the main attitude that directs future planning, intentions here will be considered to direct specific actions. Accordingly, a "plan" is composed of a sequence of conditional actions to achieve a

certain goal whereas an "intention" is the explanation behind a specific action. In other words, an intention can be thought of as a "label" to a certain action. Actions themselves can be simple atomic or complex.

Intention recognition can be seen as a substitute or complimentary to reliable and extensive communication which is a prerequisite for coordination and cooperation. If agents are able to express their intent clearly and honestly then intention recognition reduces to communication. But since not all agents are explicitly aware of their intentions or since communication can be a burden on agents (different designs, different levels of intelligence, or heterogeneous ontologies) then intention recognition becomes essential. Furthermore, it is sought to have a natural interaction in human-machine cooperation; natural to a level that it resembles human-human interaction.

Intention recognition is defined, in general terms, as the process of becoming aware of the intention of another agent. More technically, it can be defined as the problem of inferring an agent's intention through its actions and their effects in the environment. It lies accordingly in the boundary between perception and cognition. It is generally not required that the agent whose intention is being recognized has any explicit internal representation of intention or even an explicit intent. This is true as intention can be ascribed. However, agent's rationality is usually a common assumption in intention recognition.

This paper describes an intention recognition approach that relies on four-level decomposition of intentional behaviour. Figure 1 shows these levels. The first two levels constitute the classical intentional level. The desired state, either represented explicitly or implicitly, is the final result of the cognitive and planning process and is seen as its output. The actions, therefore, are selected to achieve the desired states. Here, inference is done through Dynamic Bayesian Networks

2. DYNAMIC BAYESIAN NETWORKS

Probabilistic reasoning approaches have profound advantages in application domains where uncertainty is intrinsic. Agents usually act under uncertainty as they can never have a full and accurate perception of their environments. Furthermore, rational agents do not usually further their goals in a simple if-then fashion.

Bayesian reasoning in general and Bayesian artificial intelligence in particular are based on Bayes' theorem:

$$P(h \mid e) = \frac{P(e \mid h)P(h)}{P(e)} \qquad (1)$$

It asserts that the probability of a hypothesis h conditioned upon some evidence e is equal to its

likelihood $P(e \mid h)$ times its probability prior to any evidence $P(h)$ normalized by dividing by $P(e)$ (so that the conditional probabilities of all hypothesis sum to 1)
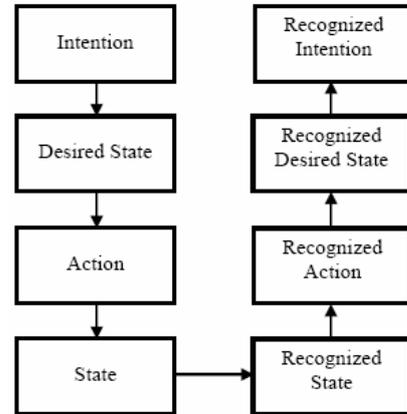
Fig. 1: Extended intention recognition scheme.

In other words and in a causal setup, it is possible to find the posterior probability of a certain cause given the likelihood of its effect. Hence, Bayes' theorem forms an excellent basis for inferring the intention from the sensed states. However, in most practical cases, the relationship between the intention and the state is not a direct and simple one. Accordingly, it becomes essential to use a representation that admits many interconnected variables and stages and makes use of conditional independence of variables. A suitable representation is Bayesian Network (BN). A BN is a directed acyclic graph encoding assumptions of conditional independence. Nodes in BN represent stochastic variables whereas arcs represent causal dependence. Associated with each node is a specification of the distribution of its variable conditioned on its predecessors in the graph. Such a network defines a joint (conditional) probability distribution (CPD); that is the probability of an assignment to the stochastic variables is given by the product of the probabilities of each node conditioned on the value of its predecessors according to the assignment. So, to define a BN, one needs to specify the structure of the network, to specify the conditional probability distribution and finally to specify the prior probability distribution of the top nodes.

In using BNs for intention recognition, intentions are represented by top nodes. The result of the intentional deliberations is a desired state. Actions and changes in states that follow from these intentions and desired states are represented by nodes below connected causally to intention nodes. Observable effects (states as measured by sensors) occupy usually bottom nodes. The prior probability distribution of the intention nodes reflects the problem context or the intending agent mental state. Intentions can, thus, be recognized by a bottom-up inference. Here, only the values of the observed

nodes are used to compute the posterior probability distribution for intention nodes.

To take into account the temporal dynamics of the "intention-activity-state" process, the Dynamic Bayesian Network (DBN) is used. DBN is a special BN architecture for representing the evolution of variables over time. It consists of a sequence of time-slices where each time-slice contains a set of variables representing the state at the current time. A time-slice is in itself a BN, with the same network structure replicated at each time-slice. The temporal dynamics of the process is encoded via the network links from one time-slice to the next. To construct a DBN, one must specify three kinds of information: the prior distribution over the variables, the transition model (from one time-slice to the next), and the conditional probability distribution.

As human intentions and actions depend on the state of the environment, cycles appear in the network. However, as a DBN is acyclic graph by definition, one cannot add cycles without expecting difficulties. Actually, all known exact inference algorithms assume "acyclicity". In this work, a novel method of eliminating cycles by time delay is proposed. As the cycles arise due to the feedback of sensed states to the intention and action nodes, one can eliminate the cycles by feeding back the sensed states from a previous time slice instead of the current one. This is quite logical as the sensed states are not instantaneously attainable but become available only after some time delay. Figure 2 shows the modified scheme to cope with the cycles. This way, the cycles are removed from the DBN without any approximation or transformation. Even in cases when cycles are instantaneous, this approach serves as an excellent approximation.
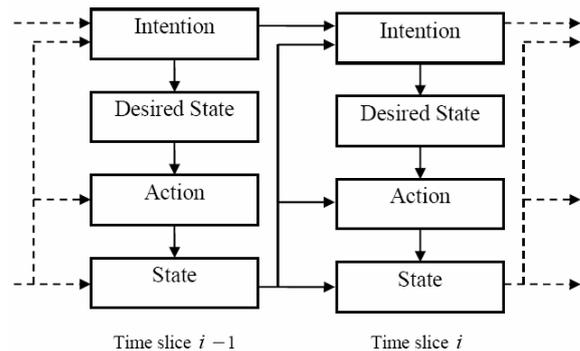
Fig. 2: DBN corresponding to human model with time-delay (no cycles)

3. COMPLIANT HUMAN-MACHINE INTERACTION

There has been a lot of methodologies presented in the literature dealing with the issues of human-machine interaction (HMI), human-computer interaction (HCI), and human-robot interaction (HRI). It is not the scope of this current work to review these methodologies neither to establish new foundations for this important field. The ultimate goal of this work is to incorporate the intention recognition module into human-machine interaction to achieve compliant and intelligent interaction. For this, an architecture for compliant human-machine interaction is proposed in Figure 3. Classical modules as human-machine interface and supervision and control can be integrated in this architecture. Although one of the goals of intention recognition is to minimize the need for traditional interfaces, it is still possible to augment certain interaction aspects with visual and auditory interfaces. General aspects of human-machine interfaces are discussed in (Johannsen, 2004). On the other hand, human supervisory control can be seen as a general framework for delegating originally human tasks to the machine. Although the intention recognition module lies roughly in this category, it is distinguished as a separate module as it relies on a different taxonomy than that detailed by Sheridan (Sheridan, 1992). Inagaki (Inagaki and Stahre, 2004) gives an interesting overview of supervisory control.

In the proposed architecture, the processes in the world are divided to human processes, machine processes, and shared processes. The effect of operating these processes can be partially or completely sensed by the human and the machine as world states. The human uses the sensed information to further her goals by selecting actions to achieve her desired states.
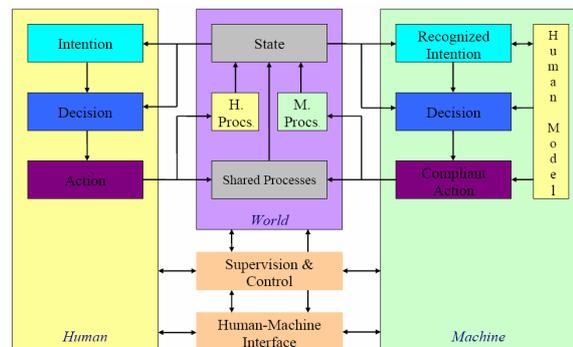
Fig. 3: Compliant Human-Machine Interaction Architecture

On the other hand, the machine which has a human-task model, employs the sensed states in recognizing the human intention and accordingly to select its action in a compliant fashion with the human intentions. The previously-mentioned division of world processes gives a great flexibility in the design as it incorporates manual mode, autonomous mode, and semi-autonomous (shared-control) mode. This flexibility widens the architecture application domain. However, and as the machine actions can be assumed to be known for the machine, the intention inference problem remains solvable despite this great flexibility. It is notable here that the relation between the machine and the world is fuzzy. The processes can be physically a part of the machine and the state

can be partially the machine state, nevertheless the distinction between the machine and the world is made to highlight the machine agency and autonomy.

4. APPLICATION EXAMPLE

Consider a human trying to manipulate a mobile robot in a given space. The robot is assumed holonomic and equipped with a laser scanner, sonars, and a camera. The human, who has access to a not-delayed robot-environment image, manipulates the robot remotely by applying forces on it through a joystick. Here, the dynamics of the robot, excited by those forces, dominates its response as it has a mass and moves on a surface with friction. In consequence, the human must be well trained to be able to manipulate it without colliding with objects or taking much time to achieve the goal. Three human limitations subdue the operation: far from perfect accelerating and braking manoeuvres, delayed human response, and inconsistent (fluctuating, spatially-confused, …) commands. It becomes then clearly advantageous that the robot recognizes the intention of the human and complies with it.

In order to simplify the discussion and the presentation, it is assumed that the space is composed of objects (existent and virtual). The motion of the robot is then described in relation to those objects. Four atomic intentions are considered: move to an object, move parallel to an object, move away from an object, and do nothing regarding an object. These intentions are atomic as they don't constitute plans and they don't require full knowledge about the environment; local sensor information suffices, namely distance and direction to object and robot velocity relative to the object. The human input (deriving the robot) is dissolved into two components (in direction of an object and orthogonal to it). Figure 4 shows the BN corresponding to this. It consists of 13 nodes corresponding to the four proposed levels:

1. Intention: intention node I to denote the four discrete intentions (G: move to an object, R: move away from an object, P: move parallel to an object, and N: do nothing regarding an object).
2. Desired state: desired velocity in the x and y directions relative to the object (D_{Vx} and D_{Vy}) and desired distance to the object D_D. The x direction is selected as the shortest line to the object.
3. Action: applied forces in the x and y directions by the human, the machine and their total (HF_x, HF_y, MF_x, MF_y, TF_x, and TF_y).
4. State: sensed velocities (V_x and V_y) and sensed distance D.

The arrows in the network correspond to causal relationships. D_{Vy}, for example, is set according to

the intention while it, at the same time, results in applying a force HF_y. This last force causes the robot to move with velocity V_y. The two directions x and y are chosen in such a way to decouple the forces. All network stochastic variables (nodes) are chosen to be discrete to simplify the inference as discussed later. The causal relationships are characterized by quantitative probabilistic models in the form of, for example

$$P(D_{Vx} = low\_positive \, / \, I = G \, \& \, D_D = medium) = 0.7$$
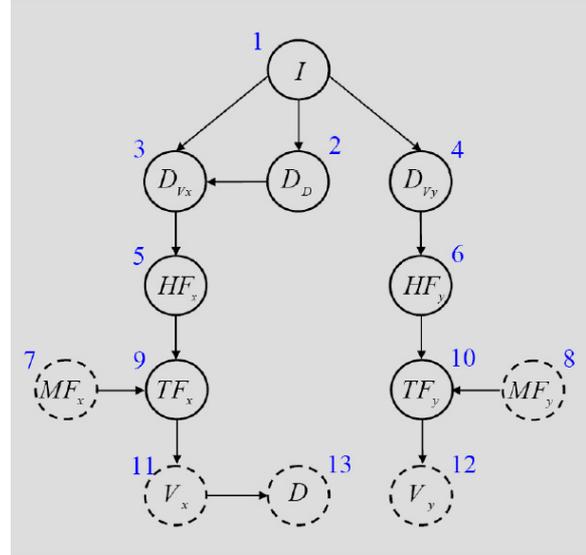
Fig. 4: BN corresponding to human – robot interaction

These relationships are gathered finally in Conditional Probability Distribution (CPD) tables. The topology of the network together with the CPDs defines the network. The network topology and its CPDs can be learned automatically; nonetheless it is quite easy to set them manually by common sense for this abridged example

The observed nodes are denoted by dashed circles while the hidden nodes are denoted with solid ones. The hidden nodes, especially the intention node, are inferred given the value of the observed nodes.

Since this is essentially a dynamic process, temporal dimension should be added to the network. Figure 5 shows the corresponding DBN (two time slices; the first is the BN while the second is shown in the figure). The first time slice is connected to the second shown in the figure by the dashed lines. Note that no cycles exist as feedback from previous time slice is used instead of the current time slice. Extra CPDs are given for nodes changed in the second time slice as for example nodes 14, 18, and 19.

Only the last level of the machine part (the action) is included. This is justified as it is assumed that the robot action is known to it, i.e. the two nodes MF_x and MF_y are observable. These nodes constitute the final result of robot decision cycle and form the

interface to the human part. Hence, the machine nodes, which lie above the action nodes, are of no interest to the inference problem. The rules governing the above nodes are given in the sequel. The networks assume that the process is a shared one where the human and robot actions can be summed. Here, the motor-schema methodology is adopted where the total action is taken to be the average of the human and robot actions. Nevertheless, it is possible to adopt any suitable arbitration method (subsumption or semi-autonomous architectures). The method adopted is then to be modeled by fixing the equivalent conditional probability distribution for TF_x and TF_y.
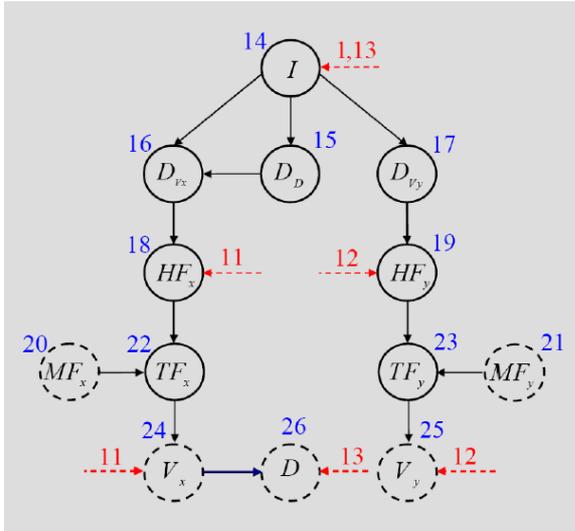
Fig. 5: DBN corresponding to human – robot interaction

For this specific example, the robot has a mass of 30 kg and a damping ratio of 15 kg/sec. Furthermore, the maximum applied force is limited to 30 N. Consequently, the following robot derivative control loop generates reasonable response:

$$MF_x = 15 * (MD_{Vx} - V_x) + 15 * MD_{Vx} \qquad (2)$$

where MF_x is the robot-applied force, MD_{Vx} is the robot-desired velocity, and V_x is the measured velocity. Force saturation is cascaded with the control law to keep the force within its maximum. The same holds for the y direction. The robot desired velocities, which are needed for equation (2), are selected in compliance with the recognized intention and based on the sensed velocity and distance:

$$MD_{Vx} = \begin{cases} 0.2 * D & I = G \\ -2 / D & I = R \\ 0 & I = P \text{ or } I = N \end{cases} \qquad (3)$$

$$MD_{Vy} = \begin{cases} 0 & I = G \text{ or } I = R \\ V_y & I = P \text{ or } I = N \end{cases} \qquad (4)$$

This simple compliant robot behavior compensates for imperfect human action. For example, it is responsible for reducing the speed when approaching an object by applying a negative force depending on the actual measured speed.

To test the validity of the proposed methods, an animation environment is built in MATLAB using its graphical user interface utility. The robot is given sensing capabilities of objects close to it and can sense its own speed. A human commands the robot by applying forces on it by moving a joystick. The inference algorithm accesses information only about the observable nodes. It employs a discretized version of these stochastic variables based on predefined discretization rules. For the inference, a smoothing engine based on a junction tree inference engine is used. This is supported in the MATLAB toolbox "BNT" developed by Murphy (Murphy, 2003).
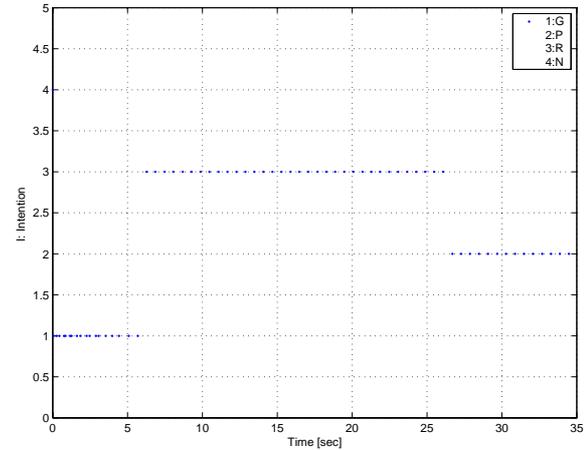
Fig. 6: Recognized Intention

An example is discussed in details to highlight the actual human-robot cooperation. The robot is located at the bottom left corner of a room and a long object covers the faraway wall of the room. The task is to move the robot toward the object (I=G), to move it parallel to the object in the right direction (I=P , V_y is negative), to move it parallel to the object in the opposite direction (I=P , V_y is positive), then finally to move it away from object to its initial location (I=R). On purpose, the human does not command the robot perfectly. Figures 6-9 show data obtained from the GUI environment. The recognized intention as shown in Figure 6 corresponds well with the actual intention. It starts with N for one sampling time, followed by G , then by P , and finally by R . The human applied force in the direction of the object is shown together with the robot's force in Figure 7. The perfect human behavior would be to accelerate at the beginning and then to decelerate in order to reach the object in a short time without colliding with the object. It is clear from the simulation that the robot complies with the human intention as it exerts the ideal force according to Equation 2. In the second stage, the robot tries to keep the distance to the object constant; this can be seen in Figure 7 with the positive force (MF_x) to compensate for the negative force applied by the

human. In the stages (G, R), the ideal thing is to minimize the forces perpendicular to the object; the effort of the robot to do this can be noticed in Figure 8. For the (P) stage, the robot tries to exert the same force as the human does; it is desired that only the human decides on the speed of motion parallel to the object. However, the robot has access only to the observable states without knowing the applied human action. So, it senses the actual velocity (V_y) and applies a force to maintain this velocity according to Equations 2 and 4. This effort can be observed as well in Figure 8.
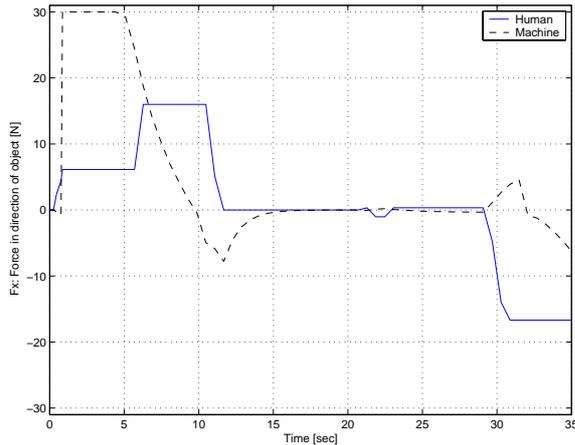
Fig. 7: Human and machine applied forces in the direction of object
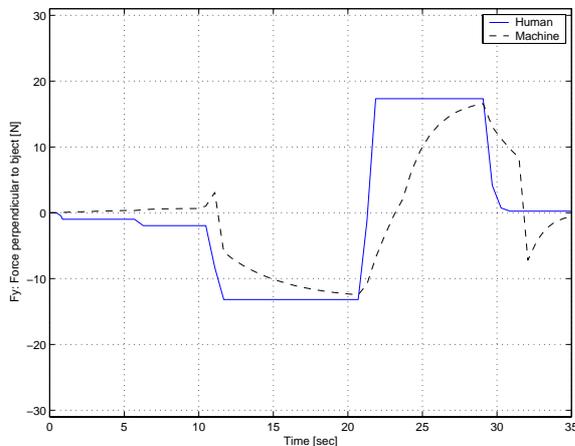
Fig. 8: Human and machine applied forces perpendicular to object

5. CONCLUSIONS

In this article, the novel idea of compliant human-machine interaction based on human intention recognition is introduced. Intention recognition and human-machine interaction modalities are introduced to achieve the goal of human-machine interaction similar to human-human interaction. A human-machine interaction architecture that enables for basing the machine action on the recognized human intention together with the state of the world is presented. The machine acts in a compliant fashion with the human intentions according to a predefined "ideal" action model while taking into account the state of the world.
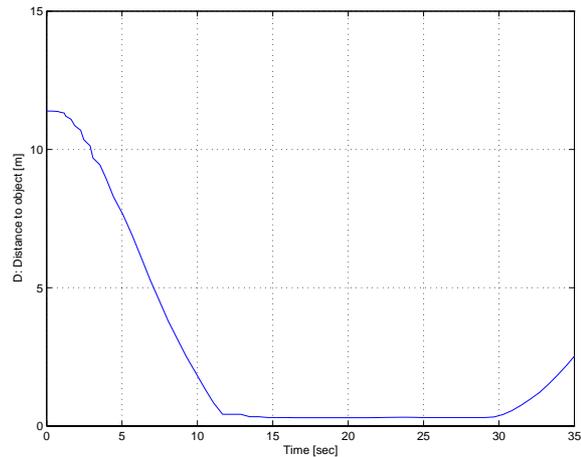
Fig. 9: Robot Distance to object

ACKNOWLEDGMENTS

REFERENCES

Bratman, M. (1999). Faces of Intention: Selected Essays on Intention and Agency, Cambridge University Press, United Kingdom.

Inagaki, T. and J. Stahre (2004). Human supervision and control in engineering and music: similarities, dissimilarities, and their implications, Proceedings of the IEEE, 92, No. 4 pp. 589-600.

Johannsen, G. (2004). Auditory displays in human-machine interfaces, Proceedings of the IEEE, 92, No. 4, pp. 742-758.

Murphy, K. (2003). Bayes Net Toolbox for MATLAB, http://www.ai.mit.edu/~murphyk/ Software/BNT/bnt.html,.

Sheridan, T (1992). Telerobotics, Automation, and Human Supervisory Control, MIT Press, Campridge, USA.

Tahboub, K. A. (2001). Natural and manmade shared-control systems: an overview, in IEEE International Conference on Robotics and Automation, Seoul, Korea, pp. 2655-2660.