Ph.D. Thesis

Giuseppe Balistreri

February, 15 2012

.... Dedica .....

Sommario

Abstract

Acknowledgments

.... Grazie .....Prof. Antonio Chella Prof. Hiroshi Ishiguro Dr. Shuichi Nishio Dr. Rosario Sorbello

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Android science is a new interdisciplinary framework useful to studying human cognition and interaction based on the premise that a very human-like robot can elicit the sort of responses people typically direct toward each other (Ishiguro (2006), MacDorman and Ishiguro (2006)).

This new robotics research area uses artificial systems, named "android", designed to be indistinguishable from humans in its external appearance and behavior.

Android are testbeds for research on the human-robot interaction, because they allow us to compare the current robot technologies and humans in a direct manner, and to verify cognitive and psychological hypotheses. Such considerations can then be applied for improving the androids.

Android science represents not only a point of connection between robotics and cognitive science, but also between appearance and behavior of interactive robots. Since we tend to anthropomorphize targets of our communication, we develop an high expectation from a humanoid robot.

This ability of the android to elicit social responses allows androids to provide a well-controlled experimental apparatus for studying human interaction and a testbed for developing theories about how neural or cognitive processes influence interaction.

Android development extends beyond the scope of engineering because, to make androids humanlike, it is necessary to investigate human interaction, and to evaluate theories of human interaction accurately, the theories need to be implemented in androids.

According to this hypothesis would be possible for an android to be used in social, psychological, cognitive and neuroscientific experiments with human participants.

Chapter 2

Android Science

The disciplines from the social and cognitive sciences potentially involved in the process of hypothesis formation and verification through human android interaction are depicted in figure 2.1.

Cognitive science employs the robot for verifying hypotheses focused on the understanding of humans (Nishio et al. (2007)).

Current robotics research uses various findings from the field of cognitive science, especially in the human-robot interaction area, trying to adopt findings from human-human interactions with robots to make robots that people can easily communicate with.

At the same time, cognitive science researchers have also begun to utilize robots. As research fields extend to more complex, higher-level human functions such as seeking the neural basis of social skills (Blakemore and Frith (2004)), expectations will rise for robots to function as easily controlled apparatuses with communicative ability.

However, the contribution from robotics to cognitive science has not been adequate because the appearance and behavior of current robots cannot be separately handled. Since traditional robots look quite mechanical and very different from human beings, the effect of their appearance may be too strong to ignore. As a result, researchers cannot clarify whether a specific finding reflects the robot's appearance, its movement, or a combination of both.

We expect that this problem can be solved using an android that has an appearance and behavior really close to humans.

The same thing is also an issue in robotics research, since it is difficult to distinguish whether the cues pertain solely to robot behaviors. An objective, quantitative means of measuring the effect of appearance is required. Androids are robots whose behavior and appearance are highly anthropomorphized. Developing androids requires contributions from both robotics and cognitive science. To realize a more human-like android, knowledge from human sciences is also necessary.

At the same time, cognitive science researchers can exploit androids for verifying hypotheses in understanding human nature. This new, bi-directional, cross-interdisciplinary research framework is called android science (Ishiguro, 2005). Under this framework, androids enable us to directly share knowledge between the development of androids in engineering and the understanding of humans in cognitive science (Figure 2.1).

•

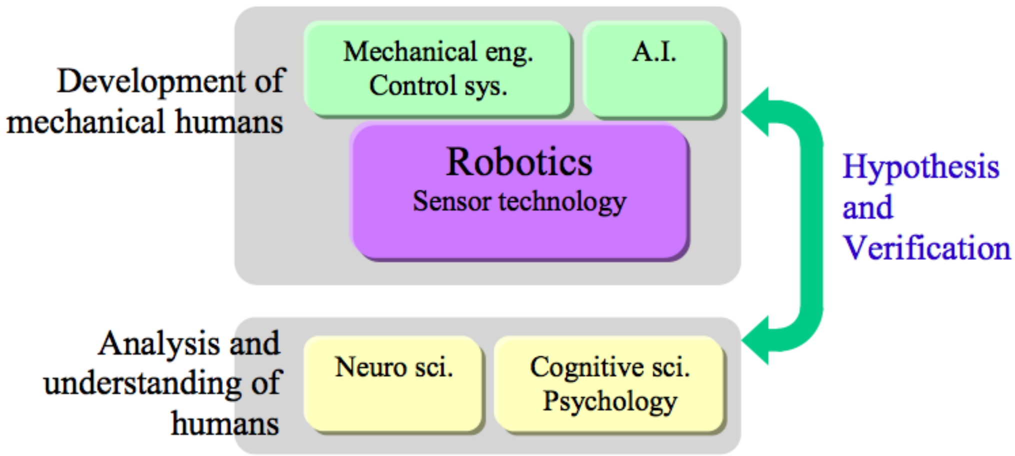

Figure 2.1: Bi-directional feedback in **Android Science**.

The major robotics issue in constructing androids is the development of human-like appearance, movements, and perception functions. On the other hand, one issue in cognitive science is "conscious and unconscious recognition". The goal of android science is to realize a human-like robot and to find the essential factors for representing human likeness. How can we define human likeness? Further, how do we perceive human likeness? It is common knowledge that humans have conscious and unconscious recognition. When we observe objects, various modules are activated in our brain. Each of them matches the input sensory data with human models, and then they affect reactions. A typical example is that even if we recognize a robot as an android, we react to it as a human. This issue is fundamental both for engineering and scientific approaches. It will be an evaluation criterion in android development and will provide cues for understanding the human brain's mechanism of recognition. So far, several androids have been developed. Repliee Q2, the latest android (Ishiguro, 2005), is shown in the middle of Figure 1. Forty-two pneumatic actuators are embedded in the android's upper torso, allowing it to move smoothly and quietly. Tactile sensors, which are also embedded under its skin, are connected to sensors in its environment, such as omnidirectional cameras, microphone arrays, and floor sensors. Using these sensory inputs, the autonomous program installed in the android can make smooth, natural interactions with people near it. Even though these androids enabled us to conduct a variety of cognitive experiments, are still quite limited. The bottleneck in interaction with human is its lack of ability to perform long-term conversation. Unfortunately, since current AI technology for developing human-like brains is limited, we cannot expect human-like conversation with robots. When meeting humanoid robots, people usually expect human-like conversation with them. However, the technology greatly lags behind this expectation. AI progress takes time, and such AI that can make human-like conversation is our final goal in robotics. To arrive at this final goal, we need to use currently available technologies and understand deeply what a human is. Our solution for this problem is to integrate android and teleoperation technologies.

2.1 Uncanny Valley

(cite: Bartneck, C.; Kanda, T.; Ishiguro, H.; Hagita, N.; , "Is The Uncanny Valley An Uncanny Cliff?," Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on , vol., no., pp.368-373, 26-29 Aug. 2007)

In the past robotics community didn't appreciate the value of building androids. In 1970, Masahiro Mori, an influential roboticist, cautioned against building robots that appear too human-like because they could be eerie or unsettling. This ability of an android to elicit human-directed responses is a phenomenon Masahiro Mori identified as the uncanny valley (Mori (1970)).

As a robot gets close to human in its outer appearance, it seems more familiar to us, until a point is reached at which subtle imperfections give a sensation of strangeness.

The uncanny valley theory was proposed originally by Mori (1970) and it's shown in figure 2.2. It hypothesizes that the more human-like robots become in appearance and motion, the more positive the humans' emotional reactions towards them become. This trend continues until a certain point is reached beyond which the emotional responses quickly become negative. As the appearance and motion become indistinguishable from humans the emotional reactions also become similar to the ones towards real humans. When the emotional reaction is plotted against the robots' level of anthropomorphism, a negative valley becomes visible (see Figure 1 picture of uncanny valley) and is commonly referred to as the uncanny valley. Movement of the robot amplifies the emotional response in comparison to static robots. A possible explanation of the uncanny phenomenon may be related to the framing theory (cite M. Minsky, "A Framework for Repre-

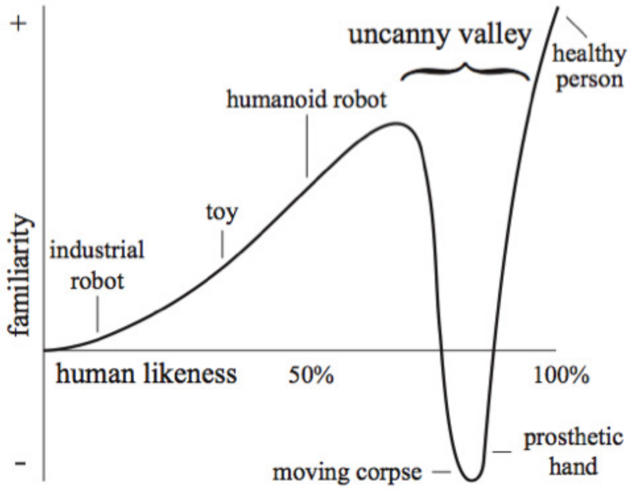

Figure 2.2: The uncanny valley

senting Knowledge," in The Psychology of Computer Vision, P. Winston, Ed. New York: McGraw-Hill, 1975.). When we encounter new situations or artifacts, we select from our memory a structure called a frame. Frames are data structures for representing stereotyped situations or artifacts. When we enter a restaurant, for example, we already have certain expectations. Attached to each frame are several kinds of information which help us in knowing how to use the frame, anticipating what will happen next and also knowing what to do when our expectations are not fulfilled. When we encounter a very machine-like robot, we select a "machine frame" and its human-like features deviate from our expectation and hence attract our attention. This deviation is usually positive since we tend to like other humans. In contrast, when we encounter an android, we select our "human frame" and its machine-like features grab our attention. However, the machine-like features are deviations that are otherwise only found in sick or injured people, which we find to be disturbing (cite: N. L. Etcoff, Survival of the prettiest : the science of beauty, 1st ed. New York: Doubleday, 1999.). In his original paper, Mori plots human likeness against (shinwa-kan), which has previously been translated to "familiarity". Familiarity depends on previous experiences and is therefore likely to change over time. Once people have been exposed to robots, they become familiar with them and the robot-associated eeriness may be eliminated (cite: M. P. Blow, K. Dautenhahn, A. Appleby, C. Nehaniv, and D. Lee, "Perception of Robot Smiles and Dimensions for Human-Robot Interaction Design," presented at 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06), Hatfield, 2006.). To that end, the uncanny valley model may only represent a short phase and hence might not deserve the attention it is receiving. We also questioned whether Mori's shinwa-kan concept might have been "lost in translation", and in consulation with several Japanese linguists, we discovered that shinwa-kan is not a commonly used word, nor does it have a direct equivalent in English. In fact, "familiarity" appeared to be the least suitable translation compared to "affinity" and in particular to "likeability". It is widely accepted that given a choice, people like familiar options because these options are known and thereby safe, compared to an unknown and thereby uncertain option. Even though people prefer the known option over the unknown option, this does not mean that they will like all the options they know. Even though people might prefer to work with a robot they know compared with a robot they do not know, they will not automatically like all the robots they know. Therefore, the more important concept is likeability, and not familiarity. Several studies have started empirical testing of the uncanny valley theory. Both Hanson (cite: D. Hanson, "Exploring the Aesthetic Range for Humanoid Robots," presented at CogSci Workshop Towards social Mechanisms of android science, Stresa, 2006.) and MacDorman (cite: K. F. MacDorman, "Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley," presented at ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science, Vancouver, 2006.) created a series of pictures by morphing a robot to a human being. This method appears useful, since it is difficult to gather enough stimuli of highly human-like robots. However, it can be very difficult, if not impossible, for the morphing algorithm to create meaningful blends. The stimuli used in both studies contain pictures in which, for example, the shoulders of the Qrio robot simply fade out. Such beings could never be created or observed in reality and it is of no surprise that these pictures have been rated as unfamiliar. This study focuses on robots and humans that can actually be observed in reality. However, it is difficult to find examples of entities in the section of the uncanny valley between the deepest dip and the human level. We are not certain if this section actually exists thereby prompting us to suggest that the uncanny valley should be considered more of a cliff than a valley, where robots strongly resembling humans could either fall from the cliff or they could be perceived as being human. We therefore included pictures of the most human-like artificial faces (computer graphics) and pictures of slightly altered humans in our study. In addition we also included pictures from real humans in our stimuli. In essence, we are approaching the uncanny valley from its far end, backwards.

2.2 Geminoid

The Geminoid is a new category of robot which has been designed to overcome the current bottleneck of robotics, which is the tight coupling between the appearance

Figure 2.3: Geminoid F and Geminoid HI-1

and the behavior of a robot.

The term "geminoid" had been coined from the Latin "geminus", which meaning is "twin" or "double", and from "oides", which indicates "similarity" or being a twin.

A geminoid is a tele-operated android which appears and behaves as an existing person, and using a computer network is possible to control the robot from anywhere. Geminoids extend the applicable field of android science allowing to focus on the human likeness and on the existence of the person, and studies about human nature in general can be performed.

Furthermore, with geminoids we can study such personal aspects as presence or personality traits, tracing their origins and implementation into robots. Figure 2.3 shows the two geminoids currently developed: on the right side of the picture there is the Geminoid HI-1, the first geminoid prototype which is copy of professor Hiroshi Ishiguro, and on the left side the Geminoid F, which is a female prototype.

2.2.1 System overview

The two developed geminoid prototypes, HI-1 and F, consists of roughly three elements: a robot, a central controlling server (called "Geminoid Server"), and a teleoperation interface (Figure 2.5).

Geminoids have the following capabilities:

- **Appearance and behavior highly similar to a living person.**

  The robotic element of the system has essentially identical structure as previous androids (Ishiguro, 2005). However, the efforts were concentrated on making a robot that appears not just to resemble a living person to be a

Figure 2.4: Dr. Hiroshi Ishiguro and his geminoid

copy of the original person. Silicone skin was molded by a cast taken from the original person; shape adjustments and skin textures were painted manually based on MRI scans and photographs. Fifty pneumatic actuators drive the robot to generate smooth and quiet movements, which are important attributes when interacting with humans. The allocations of actuators were decided so that the resulting robot can effectively show the necessary movements for human interaction and simultaneously express the original person's personality traits. Among the 50 actuators, 13 are embedded in the face, 15 in the torso, and the remaining 22 move the arms and legs. The softness of the silicone skin and the compliant nature of the pneumatic actuators also provide safety while interacting with humans. Since this prototype was aimed for interaction experiments, it lacks the capability to walk around; it always remains seated. Figure 1 shows the resulting robot (right) alongside the original person, Dr. Ishiguro (Figure 2.4).

The existence of a real person analogous to the robot enables easy comparison studies. Moreover, if a researcher is used as the original, we can expect that individual to offer meaningful insights into the experiments, which are especially important at the very first stage of a new field of study when beginning from established research methodologies.

- **Teleoperation interface (remote control).**

  Since geminoids are equipped with teleoperation functionality, they are not only driven by an autonomous program. By introducing manual control, the limitations in current AI technologies can be avoided, enabling long-

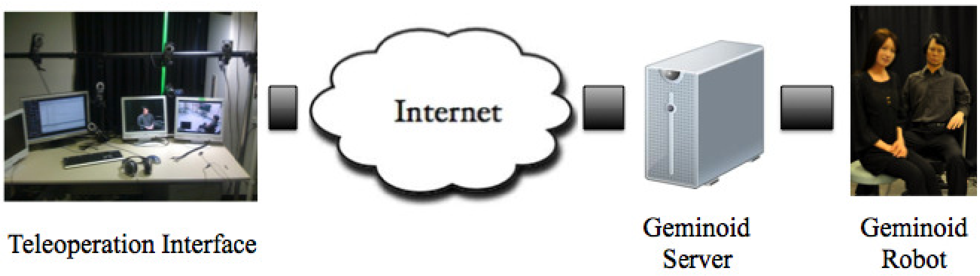

Figure 2.5: Overview of the Geminoid System

Figure 2.6: Teleoperation interface

term, intelligent conversational human-robot interaction experiments. This feature also enables various studies on human characteristics by separating "body" and "mind".

In geminoids, the operator (mind) can be easily exchanged, while the robot (body) remains the same. Also, the strength of connection, or what kind of information is transmitted between the body and mind, can be easily reconfigured. This is especially important when taking a top-down approach that adds/deletes elements from a person to discover the "critical" elements that comprise human characteristics. Before geminoids, this was impossible.

Figure 2.6 shows the teleoperation interface prototype.

Depending on the running configuration, one or two monitors show the controlled robot and its surroundings, and microphones and a headphone are used to capture and transmit utterances. The captured sounds are encoded and transmitted to the geminoid server by IP links from the interface to the robot and vice versa. The operator's facial expressions are measured by an infrared motion capturing system or by a webcam in real time, converted to motion commands, and sent to the geminoid server by the network. This enables the operator to implicitly generate suitable lip movement on the robot while speaking.

However, compared to the large number of human facial muscles used for speech, the current robot only has a limited number of actuators on its face. Also, response speed is much slower, partially due to the nature of the pneumatic actuators. Thus, simple transmission and playback of the operator's lip movement would not result in sufficient, natural robot motion. To overcome this issue, measured lip movements are currently transformed into control commands using heuristics obtained through observation of the original person's actual lip movement.

The operator can also explicitly send commands for controlling robot behaviors using a simple GUI interface. Several selected movements, such as nodding, opposing, or staring in a certain direction can be specified by a single mouse click. This relatively simple interface was prepared because, in the case of the Geminoid HI-1, the robot has 50 degrees of freedom, which makes it one of the world's most complex robots, and is basically impossible to manipulate manually in real time. A simple, intuitive interface is necessary so that the operator can concentrate on interaction and not on robot manipulation. Despite its simplicity, by cooperating with the "Geminoid Server", this interface enables the operator to generate natural human-like motions in the robot.

- **The "Geminoid Server"**

  In order to control the geminoids, a proper software system has been developed. The Geminoid family is made by several and different kind of robots. Geminoid HI-1 is very complex, has 50 axes air actuated. Geminoid F is much more simplified, having only 12 axes air actuated. The software can control also the Telenoid robot, which has a very different architecture. It has only 9 axes, and the motors are electrical.

  In order to deal with a such variety of robot architectures, we developed a new software system. This system is written in JAVA and is based on the client/server architecture. The server software is running on a PC physically close to the robot and connected with a USB interface.

  The client is a software running on a internet connected PC. The client is the software entity which selects the position of the joints of the robot. With the recent improvement of the *Geminoid Server*, different clients can be connected to the robot at the same time resulting in complex motions and behaviors. For instance, is it possible to make the robot show breathing or idle movements of the head or of the the whole body by running client which continuously sends commands to the robot.

Those idle movements can be overridden by the ones sent by the face tracking client, which starts operate when somebody sits in front of the control PC. In this case both the commands sent by the idle client and the face tracking software will try to access to the same axes placed in the head of the robot, but the one sent by the face tracking client will have an higher priority than the one sent from the idle client, and they will sent out to the robot to produce the moving.

2.2.2 The Axis Table

The "Geminoid Server" basically uses a table to store the joint values to be sent to the connected robot. The number of the rows of this table is defined by the number of the axes of the connected robot, while the columns number changes over the time, as the table works as a buffer. The server adds a new column on the left side of the table every time-step value, which is typically set to $50ms$. The values to be inserted on this column are the integer values (between 0 and 255) received by the current connected clients.

For instance, if a command like **moveaxis 2 37 5 1** is received by the server, the server will store the axis value "37" for the joint number "2" on the "Axes Table". Since there are no other attempt to use the same joint number 2 from other clients, the priority value "5" will not be used. The value "1" ending the command means that the duration of the command is just one time-step value, which roughly means only one memory cell. This scenario is depicted in figure 2.7 on the left side.

It is very common to have several clients connected at the same time, each one trying to access the same joint. In this case, the priority value sent within the command is used by the server. This case is showed in figure 2.7 on the right side, and the commands **moveaxis 2 37 5 1** and **moveaxis 2 121 0 1** are supposed to be received from the "Geminoid Server". In this case, two commands are received from the server from two different clients. Those commands are request to access the same joint of the currently connected robot, the number "2". The first command received asks the server to set the robot joint to the value "37", using a priority value of "5", while the second one asks the server to set the same joint to the value "121", using a priority value of 0. In both cases the duration of the command is one time-step. In this case, the server will discard the command with the lowest priority, and write in the table the value "37" for the axis number "2".

Some commands can be intended to last more than just one time-step. In

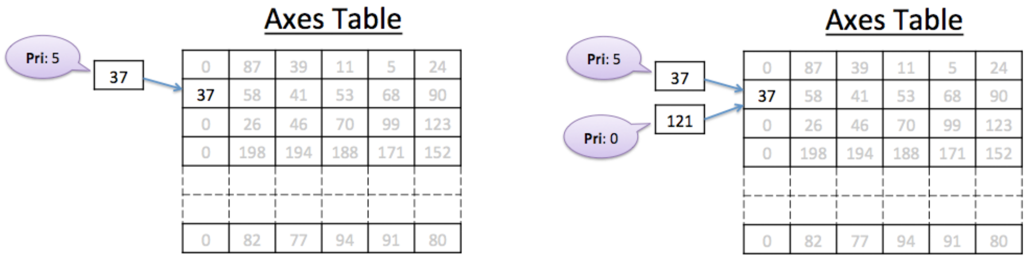

Figure 2.7: Insertion in the "Axes Table".

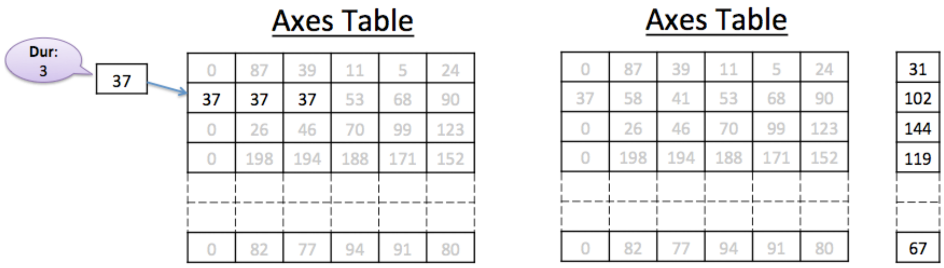

Figure 2.8: **Left side:** Duration set to "3". **Right side:** The output to the robot.

situations like this, it's possible to specify the duration of the commands as a multiple of a single time-step as the last of the four integer parameters following the "moveaxis" command. For instance if a command like **moveaxis 2 37 5 3** is received, the "Geminoid Server" will update the table like shown in figure 2.8 on the left side. The value "3" of the last integer parameter means that the duration of the action should be of three time-slots, which roughly means 150 ms. On the right side of figure 2.8 is shown the process of sending the latest commands in the table to the currently connected robot. The last column is isolated from the table, and the values saved in each cell are sent to the correspondent axis of the robot.

2.2.3 Commands accepted by the "Geminoid Server"

In this section, a list of the commands accepted from the server is presented. There are two different kind of clients: the generic client and the administration client. A generic client is any kind of software entity trying to access the axes of the robot in order to produce a motion behavior. An administration client is used for the management of the system, like set the robot to connect to or flush the table of the commands. Both kind of clients are are derived from the "Client" class, and they use an "administration protocol" or a "generic protocol" depending on their nature.

The relation between superclasses and subclasses is depicted in figure 2.9. Each

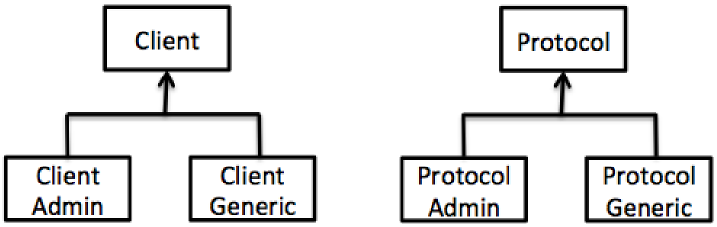

Figure 2.9: Superclasses and subclasses relation in the *Geminoid Server*.

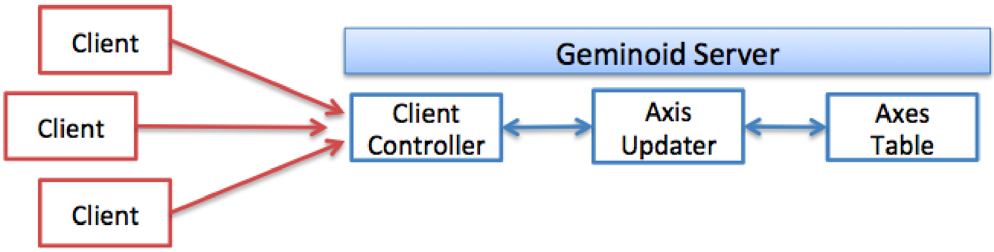

Figure 2.10: Entities interacting in a typical connection with the server.

incoming connection from a client will be managed on the server by a "Client Controller", which purpose is to keep track of the current alive connections. The commands coming from the clients are then managed by the "Axis Updater" module, which will do the priority checking of the new commands before the table insertion. If the received command has an higher priority than the existing one on the "Axes Table", an overwriting will be performed.

The figure 2.10 shows the relations between those entities.

When started, the Geminoid Server first reads a configuration file "GeminoidConfig.ini" and instantiates two software modules, "T_Starter_Admin" and "T_Starter_Generic", which will be listening for incoming connection from all the kind of clients. If a connection on the port 12000 will be detected, this means that an admin client is trying to connect, and a thread "T_AdminServer" will be started in order to manage the client. If a connection on the port 11000 will be detected, than a generic client is trying to reach the server, and a thread "T_GenericServer" will be started in order to manage the client.

The Geminoid Server also starts a software module called "GeminoidAxisUpdater", whose purpose is to manage the insertion of the incoming commands into the "AxesTable" by comparing the priority of the commands received with the priority of the ones existing on the table.

Also, a module called "CommunicationComponent" has been developed in order to send the proper set of commands to the USB connected robot.

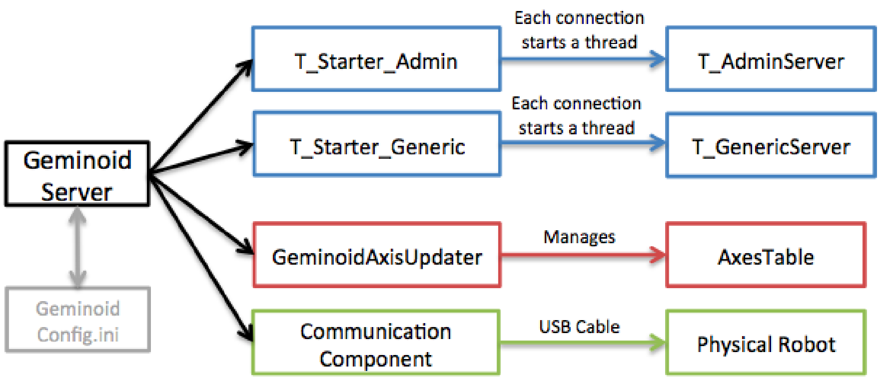

Figure 2.11: Entities interacting from the server side.

All those entities and their relations are depicted in figure 2.11.

A typical scenario consists of a running idle client, which purpose is to imitate the human breathing micro-movements, a face tracking client, which tries to emulate the facial expression of the controlling human operator, and an administration client for the management purposes.

The commands accepted for the **generic** clients are:

MOVEAXIS [AXIS_NUM] [VALUE] [PRIORITY] [DURATION]

- *MOVEAXIS* is the name of the command sent to the robot. The purpose of "moveaxis" is to change a specific axis of the robot.

- *AXIS_NUM* is an integer value that specifies the axis of the robot to be moved. The robots supported by the "Geminoid Server" have a different variety of number of joints, from 9 (Telenoid) to 50 (Geminoid HI-1), so the range of accepted values goes from 1 to the MAX_NUMBER_JOINT specified in the configuration file.

- *VALUE* is an integer value, ranging from 0 to 255. This number sets the position of the joint, from its minimum value of the angle of rotation to its maximum value.

- *PRIORITY* is a value which informs the server the value of the priority to give to the command sent. This value helps the server to understand what value has to be sent to the robot in the case that two or more clients are trying to access to the same joint. The range of possible values are from "0" (lowest priority) to 9 (highest priority). Typically, the "idle client", which emulates the breathing movements, sets the value to "0", the FaceAPI client, which tracks the facial expressions, sets the value to "5", and the administration client, which is used for maintenance tasks of the robot, and typically sets the priority value to "9".

- *DURATION* specifies the number of time-slots of validity of the current command.

ADDAXIS [AXIS_NUM] [VALUE] [DURATION]

- *ADDAXIS* is the name of the command sent to the robot. In this case is "addaxis", which meaning is to move the axis from its current position of the specified value, not from the starting ("0") position.
- *AXIS_NUM* is an integer value that specifies the axis of the robot the command refers to.
- *VALUE* is an integer value, ranging from 0 to 255. This number will be added to the current value of the axis.
- *DURATION* specifies the number of time-slots of validity of the current command.

PLAYMOTIONFILE [MOTION.TXT]

- *PLAYMOTIONFILE* is used to play a set of motion commands contained in a file on the disk. This command is useful when it is required to execute complex actions or repetitive behaviors.
- *MOTION.TXT* is the name of the file containing the list of commands to be sent to the robot.

This set of commands is extended with the ones that can be sent only from the administration client:

CONNECTROBOT and **DISCONNECTROBOT**

- are used to establish and remove the software connection between the robot and the geminoid server.

SETROBOT [GEMINOID_TYPE]

- *SETROBOT* this command specifies the physical robot connected using the USB protocol with the server.

- *GEMINOID_TYPE* specifies the type of robot. The parameters like the axis number of the robot are taken from an external configuration file. The possible choices are "GeminoidHI1", "GeminoidF" and "Telenoid" respectively if the robot used is a Geminoid HI-1, a Geminoid F or a Telenoid.

PRINTAXIS [AXIS_NUM]

- *PRINTAXIS* this command is useful to show the current queue of commands for the specifies axis saved in the "Axes Table" of the server.
- *AXIS_NUM* is the integer value which select the row of the "Axes Table" to print on the screen.

Geminoid System

Architecture of the system

The *Geminoid System* is the multithreaded priority-based JAVA-written server which has been developed in order to control the movements and the behavior of the Geminoid and Telenoid robots. In this section will be explained in detail its architecture.

- **GeminoidServer.java** This is the multithread server for the Geminoid robot. It launches a thread starter for the management of incoming generic clients *T_Starter_Client* connections and a thread starter for the management of administration clients *T_Starter_Admin* connections. Starts also an update manager *GeminoidAxisUpdater* for the command table.
- **T_Starter_Admin.java** This thread starts a new thread *T_AdminServer* each time that is received a connection on the port 4444 (which is read from the configuration file and that is used by the administration clients).
- **T_Starter_Generic.java** This thread starts a new thread *T_GenericServer* each time that is received a connection on the port 4445 (which is read from the configuration file and that is used for generic clients).
- **T_GenericServer.java** Thread implementation for the generic client's *GeminoidServer*. It uses the *ProtocolGeneric* class for making the association between the command sent from the client and correct behavior of the server.
- **T_AdminServer.java** Thread implementation for the administration of the *GeminoidServer*. Uses the *ProtocolAdministration* class for making the associations between the command sent from the client and the correct behavior of the server.

- **ProtocolGeneric.java** This class is the protocol definition used by the *GeminoidServer* for communicate with the generic client. A generic client could be a terminal client or a Face API client or an idle behavior.
- **ProtocolAdministration.java** This class is the protocol definition used by the *GeminoidServer* for communicate with an administration client. It contains a list of command for the robot maintenance.
- **Geminoid.java** This is the class that stores informations about the currently selected robot, like the type of the Geminoid or the number of its axes. It also implements some methods used for managing these informations stored in the class itself.
- **GeminoidAdminServer.java** This is the multithread administration server for the geminoid robot. It uses a configuration file called *GeminoidConfig.ini* which contains information about the port used for listening, the number of axes of the robot and so on. When it's started creates a thread which periodically checks the clients queue and moves the commands to the robot queue. Also, for every incoming connection it creates a new *GeminoidMultiServerThread* thread.
- **GeminoidAxisUpdater.java** This class updates the axis table periodically using an interval of time specified in the configuration file *GeminoidConfig.ini.*
- **GeminoidCommand.java** This is the class that implements the command sent to the server. It stores informations about the type of the command, the number of the axis involved into the command, the priority of the command which has been sent the robot and the time of the message receiving. It also implements some methods used for managing these informations stored in the class itself.
- **GeminoidCommandController.java** *GeminoidCommandController* is used for the management of the information contained into GeminoidCommand's. This class provides methods such *moveaxis*, called when a request for axis moving is received, or *addaxis* which is called when an "addaxis" command is received, and *axisUpdater* that is used for take the current column from the axis table to send to the robot.
- **Client.java** This class keeps information about the clients connected to the server. It stores for each client: *ClientID*, which is an unique ID on the server for each client connected obtained from the concatenation of the remote port and the client IP address; *ipAddress*, which is the IP address of the client

connected; *remotePort*, which is the remote port of the connected client and it's used for the creation of the *ClientID*; the *type*, which helps to understand the type of the client like an administration (if it's an administration client) or generic (if is an undefined client).

- **ClientController.java** This class manages the information about the clients currently connected to the *Geminoid Server*. It provides the add and remove and some other utility methods for the management of the list of the connected clients.

- **GeminoidConfig.ini** This is the file used to read and write the configuration values.

Introduction eye gaze article BICA

The ultimate goal of robot development from the human-robot interaction perspective is to build a robot that exhibits comprehensible behaviors and that supports a rich and multimodal interactions NISHIO and ISHIGURO (2011) Kanda et al. (2004). Recently much has been done in this direction, but in most cases the interaction with robots is efficient only with their developers or with a group of trained individuals. Furthermore, the value of the robots would strongly increase if ordinary people can accepted them as a social presence.

If a robot looks like a human and displays natural human-robot interactions, an ordinary human can naturally communicate with the robot as if he were communicating with another human. One approach for discover the principles of giving to a robot a human-like presence is to build a robot system that can sense and behave like a human. The results of this research can give us the principles underlying the human-like existence.

The degree of how much human-like nature and how much perception are needed in order to perform a natural human-robot interaction have not been revealed yet. The robot's appearance must be sufficiently anthropomorphic to elicit natural reactions from people interacting with it Shimada et al. (2006). A robot which realizes a very human-like appearance is called an "android". However, the android system of prof. Hiroshi Ishiguro Nishio et al. (2007) is not yet sufficient yet as to realize a natural human-robot interaction, because the android's perceptional functions are not implemented, or are substituted by a human controller. Is possible to improve the communication abilities of a robot using built-in sensors, but they have many limitations as to range and resolution. Some perception functions can be provided to the robot by embedding sensors in the environment Morishita et al. (2003), Mori et al. (2005), Ishiguro (1997) and creating a network of sensors (like cameras) that can overcome the limitations of the built-in ones. The resulting architecture of the system is different from the human perception system, but the modalities are not important for a natural communication. For example, humans can easily detect the position of someone else by using vision (eyes) or audio sensing (ears), but in terms of interaction the most useful information is the position of the other person. A robot can obtain the position of someone else by using sensors embedded in its surrounding environment, but it can use the information obtained as if it comes from a complex vision system like the human one.

Concerning the interaction with the robot, if we think about it as a social entity, we need to clearly define the communication skills necessary for robot to be integrated in daily life, in order to discover the principles underlying natural interaction among humans, and to establish a methodology for the development of expressive humanoids robots. Also, another important factor for a social robot is affinity with humans. If the robot is a good partner in the human-robot interaction, the value of affinity with the humans can be increased. So there is a need to understand how the interaction can be improved Takano et al. (2009).

A lot of work has been addressed by several researchers on the robot side of interaction, in particular on receiving inputs from human like speech recognition, computer vision, etc.) but, in contrast no important improvements have been done on how a robot should give feed-back to its user. There is a necessity of a transparent interface that regular people can easily interpret Bruce et al. (2002).

In the human context, the interpretation of other's behavior is a complex and at the same time efficient task. We follow social rules, and if someone doesn't behave accordingly, we have an unpleasant and annoying feeling. So, in order to be accepted in the human society, robots need to behave in ways that are socially correct. Supporting the hypothesis that face-to-face interaction is the best model of interface, we want to leverage people's ability to recognize the subtleties in eye focusing as a feedback, making the conversation with the robot richer and more effective and at the same time discovering which parameters are most significant and useful for human-robot interaction.

This paper reports the development of a communication system that integrates sensors embedded in the environment with an android.

2.3 Robot's gaze and eye contact

One of the most effective way of controlling human communication is eye contact. For example, we can start a conversation after establishing this kind of contact, or we can infer if our conversational partner is paying attention to us from his gaze. This kind of communication is called meta-communication. Recently, several robots have been proposed that utilize gaze for meta-communication: ROBITA Matsusaka et al. (1999) turns to the specific person speaking at the moment in a group conversation, Robovie Kanda et al. (2002) and COG Brooks et al. (1999) are similar examples. A measurement of the effects of the gaze behavior on believability has been performed in Poel et al. (2009). Eye contact is a phenomenon that occurs when two people cross their gaze, and since we perceive eye contact clearly, eye contact has a stronger meta-communication capability than just the gaze.

But making eye contact is not only about looking at each other, because both parties need to be aware of being watched by the other. This can be achieved with the eye focusing, which basically means to create the eye convergence mechanism on what is the object of the attention. Social robots, as well as service robots, are expected to behave in a way similar to humans, such as talking and listening. For those tasks, robots requires detailed information about the people they are interacting with. To establish an effective eye contact, the robot needs to know the exact position of the human head of the conversational partner, as people are known to be highly sensitive in distinguishing gaze directions and identifying eye contacts.

2.4 The Geminoid System

This section briefly describes our robot system used for eye-contact in a sensor network study.

A Geminoid is a robot whose purpose is to duplicate a living person. Geminoid HI-1 was developed to closely resemble the outer appearance of its creator Prof. Hiroshi Ishiguro. The term "Geminoid" Becker-Asano et al. (2010) is derived from the latin word "geminus" meaning twin and "-oides" meaning similarity. In contrast to humanoid robots Kanda et al. (2004) which are similarly designed to let people associate them with humans, the outer appearances of android robots such as Repliee R1 Minato et al. (2004), Repliee Q2 Shimada et al. (2006) or Geminoid HI-1 even feature artificial skin and hair, and they are modeled to finest detail in the aim to make them indistinguishable from real humans at first sight. With these androids it is possible to pursue research in the field of "Android Science" Ishiguro (2007), in which these special robots are seen as "a key testing ground for social, cognitive, and neuroscientific theories" MacDorman and Ishiguro (2006). The android is 140 cm tall, sits in a chair (it cannot stand) and has 50 DOFs. Its face has 13 DOFs, which gives it natural facial expressions. Figure 2.12 on the left side shows the Geminoid HI-1 and its real counterpart.

The robot is equipped with teleoperation functionality. In this way we can avoid the current limitations in AI technologies. Figure 2.12 on the right side shows the teleoperation interface. Two monitors show the controlled robot and its surroundings, and microphones and a headphone are used to capture and transmit

Figure 2.12: **Left side**: Geminoid HI-1 with its creator Hiroshi Ishiguro. **Right side**: control room for the Geminoid HI-1. Only the camera between the two monitors on the right side is used, the others five cameras are for the motion capture system (not used in this experiment).

utterances. The captured sound from the environment around the robot are transmitted to the operator, and vice-versa. A webcam points the face of the operator, and its video stream is continuously analyzed for the operator's head orientation and mouth movements using the "FaceAPI" software (**Citare FaceAPI**) The acquired head motion data is used to drive Geminoid's head orientation in the space and the mouth movements (opening and closing the lower jaw). The operator's voice is synchronized with the lip movements by delaying it for approximately 0.5 seconds such that the robot appears to speak by itself. Furthermore, a trained operator could manage to turn the Geminoid's head as if it were looking at a specific conversation partner. In parallel to these movements a separate software module continuously triggers small movements in Geminoid HI-1's face opening and closing its eye lids and moving its cheeks slightly up and down from time to time. The operator's mouth movements as well as the movements of the head are captured. The robot has been designed to work also with tactile and floor sensors but for this study only cameras have been used.

2.5 Related work

Recently, several robots have been proposed that utilize gaze for meta-communication. ROBITA turns to the specific person speaking at the moment in a group conversation. With Robovie Imai et al. (2002) has been investigated the relationship between a robot's head orientation and its gaze in a Robot Mediated Round Table (RM-RT) experimental setup. A similar task has been addressed with Cog, but to our knowledge, none seems to address the problem using an android robot.

We believe that use an android robot can give us a more detailed knowledge about the “uncanny valley" hypothesis.

2.6 Multicamera System

This section describes the idea behind the human head detection and tracking using a network of cameras embedded in the environment surrounding the android.

This is basically a slightly modified implementation of the work of Matsumoto et al. (2010). The original work is a robust implementation of an algorithm able to track people in an open space with various lighting conditions, using a combination of multi-camera which can track human heads and and laser rangefinders (LRF) which can perform stable tracking.

With respect to the original formulation of the algorithm, we made some adjustments due to the different environment in which the geminoid is placed. There is a regular and constant artificial light in the room provided by neon tubes, and there are no windows. Moreover, the surface of the room is smaller (approximately 35 square meters) than the typical environment in which the work of Matsumoto is intended to be used.

We decided to use only cameras for the tracking system and after an initial calibration the system was able to run without any changes of the parameters.

2.6.1 Distributed PF

We employ the PF Matsumoto et al. (2010) for tracking the human head. This algorithm estimates a posterior $p(\mathbf{X}_t|Z_t)$ with random sampling and its evaluation from likelihood $p(\mathbf{Z}_t|\mathbf{X}_t)$.

The reason to use the PF is that $\mathbf{X}_t$ (object state at time t) and $\mathbf{Z}_t$ (image feature at time t) have too wide space to estimate posteriors. The PF approximate the posterior as sample set $\left\{\mathbf{s}_t^{(n)}, \pi_t^{(n)}, n = 1, ..., N\right\}$ where $\mathbf{s}_t^{(n)} \in \mathbf{X}_t$ represents hypothesis of target, $\pi_t^{(n)} = p\left(\mathbf{Z}_t|\mathbf{X}_t = \mathbf{s}_t^{(n)}\right)$ represents weight of hypothesis.

The hypothesis $\mathbf{s}_t^{(n)}$ is generated according to prior $p\left(\mathbf{X}_t|\mathbf{Z}_{t-1}\right)$; previous samples-sets invent new samples with random sampling.

A generation of hypotheses $\mathbf{s}_t^{(n)} = (X, Y, Z)$ in 3D space is done in the same manner as PF. That is, these hypotheses are generated by random sampling according to previous sample-set $\left\{\mathbf{s}_{t-1}^{(n)}\pi_{t-1}^{(n)}\right\}$ with dynamic model.

We employ Gaussian G(•) for random sampling as

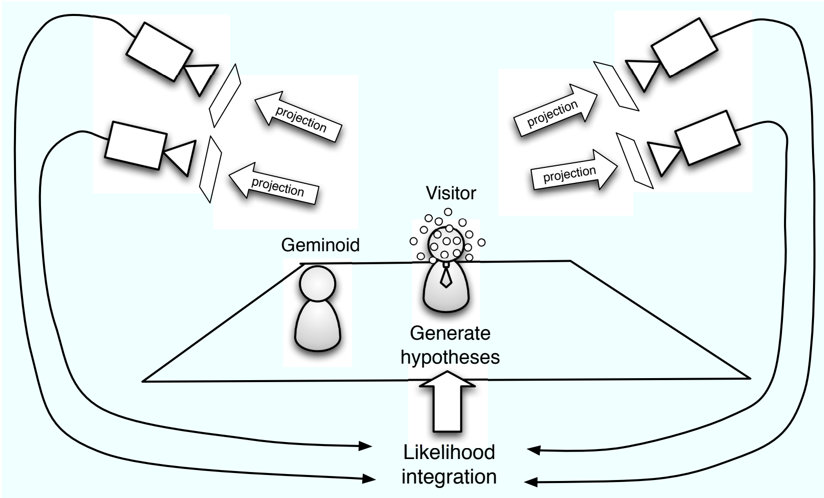

Figure 2.13: The distributed PF system

$s_t^{(n)} = G\left(s_t'^{(n)}\right)$ (2.1)

$s_t'^{(n)} = R\left(\left\{s_{t-1}^{(j)}, \pi_{t-1}^{(j)}\right\}\right)$ (2.2)

where R(•) represents a function that select temporary hypothesis $\mathbf{s}_t'^{(n)}$ from previous hypotheses $\mathbf{s}_{t-1}^{(j)}$ by ratio of weights $\pi_{t-1}^{(j)}$, and new hypothesis $\mathbf{s}_t^{(n)}$ is generated by Gaussian with the temporal sample $\mathbf{s}_t'^{(n)}$ as $G\left(\mathbf{s}_t'^{(n)}\right)$.

The algorithm allows to track multiple people generating hypotheses for multiple people as $\mathbf{s}_{m,t}^{(n)}$, where m is the person's number. Similarly, a weight is represented as $\pi_{m,t}^{(n)}$. Then, the human head position $\mathbf{P}_{m,t}$ is computed as

$\mathbf{P}_{m,t} = \sum_n \mathbf{s}_{m,t}^{(n)} \pi_{m,t}^{(n)}$ (2.3)

By repeating these steps of hypotheses generation (random sampling), 2D fitness value evaluation and 3D fitness value estimation, human head tracking is performed (Fig. 2.13).

2.6.2 Multicamera System for Head Tracking with Distributed PF

In order to estimate the likelihood of the distributed PF, all the hypotheses are projected onto 2D image taken by camera i for the evaluation of the 2D fitness value $\pi_{i,t}(\mathbf{a})$, where **a** means 3D position.

An integration of the 2D fitness values is then performed for computing 3D fitness values $\pi_{m,t}^{(n)}$, basing on the weights of 3D hypothesis.

The 3D fitness values can be calculated using the equation below,

$\pi_{m,t}^{(n)} = \sum_{i,j,i \neq j} \pi_{i,t}(\mathbf{s}_{m,t}^{(n)}) \pi_{j,t}(\mathbf{s}_{m,t}^{(n)})$ (2.4)

defined in Matsumoto et al. (2010).

The model of human head is spheroid as shown in the left side of figure 2.14. We use two constant values, l_0 and l_1, for determining the size of the spheroid.

By considering the projected 3D object specified by $\mathbf{s}_{m,t}^{(n)}$ and the related image features, is possible to evaluate 2D fitness value.

The projected spheroid onto image place will result in an ellipse like shown in figure 2.14 on the left side shows the ellipse drawn. The fitness value is obtained by evaluating the edges on the ellipse using a Sobel filtering.

Simple evaluation of edge strength on the ellipse might be affected by background edges.

For improving the sensitivity to human head, the fitness is estimated by inner product between normal vector $\mathbf{N}_\mu$ of the ellipse, which is normalized, and the gradient vector of edge $\mathbf{E}_\mu$ at μth sampling point as shown in the right side of figure 2.14, which is not normalized. The fitness is computed by the following formula:

$\pi_{i,t}\left(\mathbf{s}_{m,t}^{(n)}\right) = \sum_\mu \mathbf{N}_\mu \bullet \mathbf{E}_\mu$ (2.5)

The value $\pi_{i,t}\left(\mathbf{s}_{m,t}^{(n)}\right)$ becomes maximum when edge vector $\mathbf{E}_\mu$ and normal vector $\mathbf{N}_\mu$ have the same direction.

The sampling points are illustrated in the right side of figure 2.14, where we will not sample around the bottom of ellipse, because this area may be located on the body (Fig. 2.14).

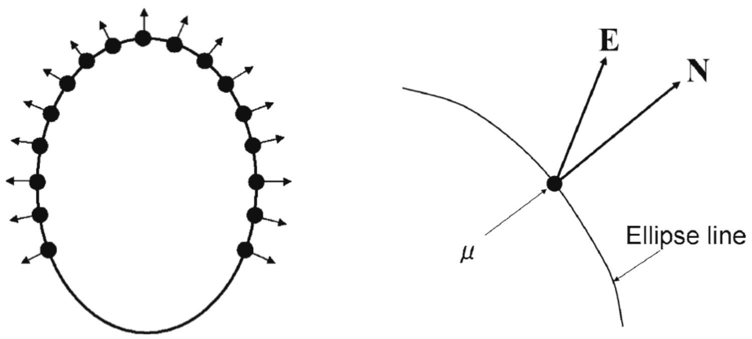

Figure 2.14: The normals on the ellipse on the left side and the parts of the normal on the right side.

2.6.3 Hardware and Software Configuration

The system implemented wants to achieve the ocular convergence on the Geminoid robot. Using a set of cameras is possible to verify a set of hypothesis generated from a PC called “master PC". In the configuration tested we prepared four cameras each one placed in one of the four corners of the Geminoid’s room. Those cameras are placed at an height of around 2,5 meters from the floor, and the wide lens used cover the area in front and on the right and left side of the robot. From the computational side, the adopted configuration was composed of a total of three PC’s: one “master PC", and two “slave PC". Each “slave PC" runs two instances of a developed client. Each client is connected to one of the cameras placed in the room. All the computers used are connected in a TCP/IP network configuration.

We used machines equipped with Intel i7 processors and 8Gb of RAM. The software environment was Windows 7 and the OpenCV libraries for the image processing.

We placed the origin of the coordinates system in the bottom left angle from the entrance door of the room.

The estimation of the head position of humans subjects around the robot is done in real time. Basically, a set of hypotheses is generated, and those hypotheses are broadcasted to each slave client. Then each client verifies the set of points received, and sends its answer back to the master PC, so that a new set of hypotheses can be generated.

Figure 2.15 shows the sequence of tasks for the master and the client PC.

Below, the sequence of steps is explained more in detail:

1. The master PC generates a set of points in the 3D coordinate system of the room, following a gaussian distribution in the 3D space. Each point is the

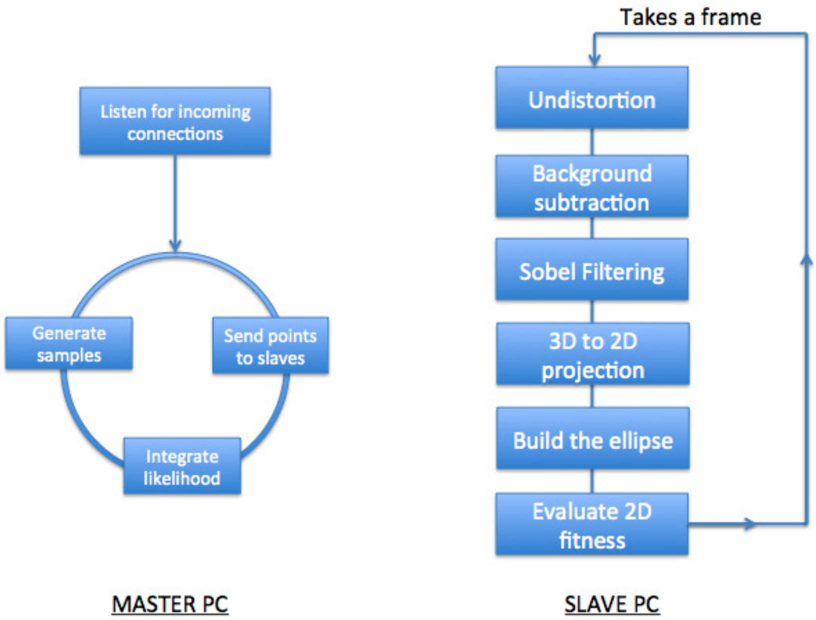

Figure 2.15: Sequence of tasks on the master PC and on the slave PC.

center of the ellipsoid representing the estimated position of the human's head. This generation considers the contribution from the previous set of particles, or, at the startup of the system, we place the center of the points in the entrance door where the people are supposed to enter the room.

2. The set of particles generated is broadcast to all the clients connected through the wired TCP/IP network. The typical loop of tasks for a client once a set of points has been received is:

   - A projection of all the points in the received set from 3D coordinate system to the 2D coordinate system of the current camera;

   - An edge detection of the current frame using the Sobel filtering. The resulting image only shows the edges of objects and persons the scene.

   - Around each 2D projected point an ellipse is built, according to parameters $l1$ and $l2$ chosen, so that the ellipse represents the human's head proportions.

   - a 2D fitness value for the current ellipse is computed. The fitness is estimated by inner product between the current normal vector of the ellipse, and the μth sampling point as shown in (2.5).

3. Each client sends back to the "master PC" the 2D fitness values obtained for each ellipse, and those values are integrated for estimating 3D fitness values

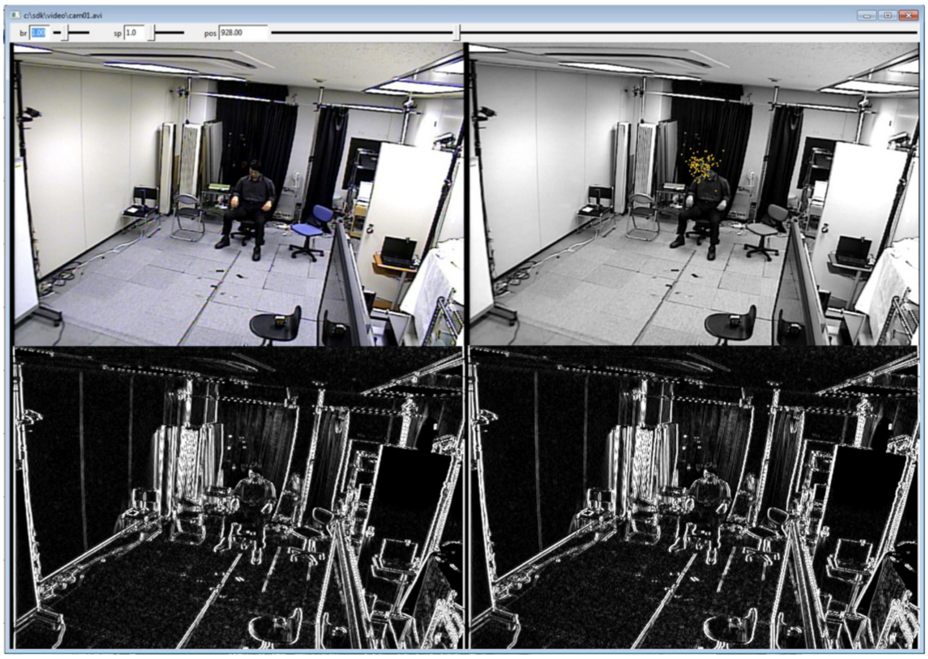

Figure 2.16: Screenshot of the system running on a slave PC.

according to the formula in (2.4). The point with the highest 3D fitness value is used as center for the generation of the new set of 3D points with the gaussian distribution.

By iterating this sequence of steps, real time tracking of the human head is performed.

Figure 2.16 shows an onverview of the software running on a slave PC which is connected to a camera placed in front of the robot.

The main window is split in four boxes. The one on the top left shows the view coming from the connected camera. The ones in the bottom part of the window show the result after the application of the Sobel filtering for the x and y derivative. The frame on the top right shows the condensation of point placed on the head of the robot, which is an exact copy in the shape of a human head.

2.7 Expected Results and Improvements in the Control System

The solution previously shown adds an efficient perceptional function to the human-like appearance of the android, which is useful for communicating through the typical human communication channels. Using the cameras placed in the four corners of the room is possible to cover the area in front of the robot. In this way we can obtain information about the position of the human subject when someone is very near in front of the robot (short range interaction) or when someone is moving in the room (long range interaction). Figure 2.17 shows the map of the room and pictures of the setup.

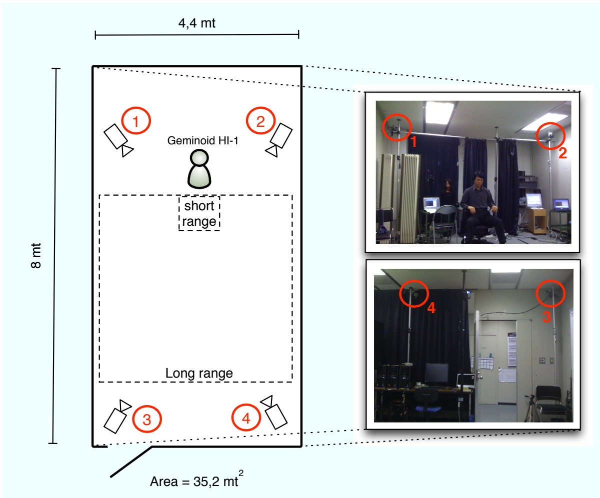

Figure 2.17: On the **left side**: map of the Geminoid room from the top showing the cameras and approximately the two ranges of interaction, the short one and the long one. On the **right side**: pictures of the setup.

2.7.1 Short Range Interaction with the Android Compared to Humans

A possible way to evaluate the human-robot interaction would be to couple non-conscious human responses together with a complementary source of information such as a questionnaire. The gaze behavior in the human-robot interaction can later be compared to the gaze behavior in human-human interaction, which has been widely studied in psychology and in cognitive sciences. A typical behavior is to establish an eye contact looking into the interlocutor's right eye with the right eye. This contact is often lost, especially when one of the conversational partners is thinking. This behavior has been explained by three main theories:

Arousal reduction theory: during a face to face conversation, arousal is highest when a person makes eye contact. As a natural response, we tend to break eye contact to reduce our arousal and concentrate on the communication Argyle and Cook (1976).

The different cortical activation hypothesis: according to this theory, the brain activation caused by thinking tasks leads individuals to move their gaze away from the central visual field Previc and SHANNON (1997).

The social signal theory: this theory gives the meaning of social signal to gaze behavior, so we break eye contact to inform others that we are thinking.

In order to resemble the human behavior in a face-to-face conversation, the android should be able to establish, maintain, break and recover an eye contact with his human conversational partner.

Our hypothesis is that if we allow the Geminoid to produce the same kinds of eye movements following the same social rules of humans, subjects will consider the robotic interlocutor as if it were a person, or at least a social agent. We believe that eye movements act as signals about whether the subject is, for example, thinking or listening, and an android must use this kind of non-verbal communication in order to increase its human-likeness.

To test the validity of this speculation, it is needed to consider the duration and the timing, together with the focus of gaze. The same kind of conclusions have been reached in MacDorman et al. (2005).

2.7.2 Long Range Interaction with the Android Compared to Humans

In Chikaraishi et al. (2008) has been observed that in a scenario consisting of a quietly sitting person (subject B) and another person free to act in a 2.0 m x 2.0 m area (subject A), the typical behaviors of B were to:

- look toward the same direction that A is looking at;
- look toward subject A a few times with only eye movements and as few body movements as possible;
- look at A a few times;
- look toward the front so as not to see subject A;
- look down to the ground;
- keep looking at subject A (following him).

In order to achieve a human-like nature in android communication, it is necessary to implement abilities equivalent to those observed from humans. Results from experiments like the one in Chikaraishi et al. (2008) suggest that reactions for a human-like presence are efficient for achieving the subjective human-like nature in android communication, and the minimum required abilities to exercise and perceive for achieving a human-like nature, like a boundary condition of natural interaction, should indicate the principles of natural communication. The system described in this paper will allow us to tackle issues focusing on appearance and perception, although such studies have not started yet. The system described can be a test bed for cognitive science, and some research approaches in cognitive sciences have used robots for experiments.

2.8 Conclusions and Future Works

This section proposed a hypothesis for the improvement of the human likeness of an android by establishing, breaking, and recovering the eye contact with a human interlocutor.

Those social signals can be used in both face-to-face interactions and long range interactions, and they can reinforce our expectation of androids as a responsive agent.

However, this study is only preliminary and a more comprehensive one is required to contribute to the study of android science and human nature. Our next work will include further investigation on the effect of the gaze for an android in order to confirm the psychological effects.

Chapter 3

Telenoid

Within the last decade, the development of very human-like anthropomorphic robots, which are at first glance indistinguishable from real humans and are often referred to as “androids”, has become feasible. Such androids have been expected to aid in the understanding of areas of human cognition that could not have been tested or clarified until now.

Very human-like androids in the past, such as the Geminoid HI-1 and Geminoid F androids developed by our research group, resembled real persons and were intended to convey the feeling that those specific individuals were present at the robot’s location. For example, a person facing such android feels the presence of the actual operator, and, when using the tele-operation system we developed, reacts to the android as if the operator was really there Nishio et al. (2007).

Telenoid R1 was designed to appear and behave like a minimalistic human; at first glance, one can easily recognize that the Telenoid resembles a human, but it can be interpreted as being either male or female, old or young. Due to this minimal design, the Telenoid allows people to feel as if a far-away acquaintance was close to them. In this paper we want to investigate ordinary people’s natural reactions and impressions outside of the laboratory in order to check that the concept for the Telenoid works.

Laboratory interactions are rather artificial in nature, because the situational context influences the participants’ expectations and attitudes Bartneck et al. (2007). Experimental laboratories are perfectly controlled environments. Therefore, results obtained within such environment can be very useful from a scientific perspective. But data regarding people’s natural impressions or reaction toward androids cannot be obtained easily in such environment. We think that the field environment - although uncontrolled - must also be an important source for obtaining knowledge toward the further development of androids.

In this paper we describe two field tests using the Telenoid. These two field tests were conducted in order to acquire insight into two questions, as follows: (1)

whether minimal humans such as the Telenoid will be a next tele-communication media and (2) whether the Telenoid R1 can be accepted by ordinary people.

3.1 Related work

Geminoid HI-1 was developed to closely resemble the outer appearance of its creator, Prof. Hiroshi Ishiguro. In contrast to typical humanoid robots Kanda et al. (2004), which are designed with a human-like shape or features in order to allow people to associate the robots with humans, the outer appearances of androids such as Repliee R1 Minato et al. (2004), Repliee Q2 Shimada et al. (2006), or Geminoid HI-1 Nishio et al. (2007) even feature artificial skin and hair, and they are modeled to the finest detail in the aim of make them indistinguishable from real humans at first sight. With these androids it is possible to pursue research in the field of Android Science Ishiguro (2007), in which these special robots are seen as "a key testing ground for social, cognitive, and neuroscientific theories" MacDorman and Ishiguro (2006).

The effects of an android's anthropomorphic appearance and body movements have so far mainly been investigated in a number of empirical studies conducted within laboratory environments. Minato et al. used the android Repliee R1, for example, to investigate the hypothesis that the uncanny feeling diminishes together with increased complexity of the android's behavior. This hypothesis was supported by a number of subsequent laboratory studies, e.g. Shimada et al. (2006), but investigating this question is still a prime motivation underlying android science research.

3.2 Telenoid

3.2.1 Specification and Teleoperation System

The Telenoid has nine degrees of freedom, or DOFs (by contrast HI-1 has fifty DOFs and F has twelve DOFs). Specifically, the provided DOFs allow independent horizontal motion for the left and the right eye, and synchronized vertical motion for both eyes, opening and closing of the mouth, yaw, pitch and roll rotations for the neck, as well as motion for the right hand and left hand. The Telenoid's length is eighty centimeters, and its weight is about six kilograms. The covering skin is fashioned from silicon and it feels pleasant and similar to human skin when touched.

The operator's face directions, mouth movements and facial expressions are captured by a face recognition system. These face tracking results are used to create commands which are sent to a server via TCP/IP. Therefore, if an internet connection is available, the Telenoid can be operated from anywhere in the world. The video stream for the face recognition system is obtained using a web camera attached to a laptop. Some specific behaviors, such as “bye-bye”, “happy” or “hug”, can be controlled by GUI buttons on the display. Some spontaneous behaviors, such as breathing, are generated automatically to create the sense that the android is alive. Expression of the breathing motion is for example obtained with a slight and periodical motion of both hands. Basically, the tele-operation system requires only a single laptop.

3.2.2 Design Concept

The aim for the Telenoid was to create a minimal human, as such design might allow any person to transfer its own presence to distant locations.

In order to achieve this purpose, the following three requirements are necessary: (1) omni-human likeness, (2) tactileness, (3) mobility. The “omni-human likeness” allows people to feel as if a far-away acquaintance was close to them. The “tactileness” and the “mobility” facilitate the interaction with the Telenoid.

The design concept used for Geminoid HI-1 and F is almost opposite to that one used for the Telenoid. Both of them have specific characteristics. For Geminoid HI-1 or F such specific features are important to convey the feeling of the intended actual human’s presence. But if an unsuitable operator controls a Geminoid these specific features might negatively affect the interaction. Besides that, Geminoids are hard to move because of their weight and size.

There are several robots that have some points in common with the Telenoid. For instance a teddy-bear-style robot, the IP RobotPHONE, aims to achieve a telepresence communication Sekiguchi et al. (2001). However the outer appearance of the robot shows very specific characteristic, and the forms of interaction used by people might be affected by this design. To create the design of a minimal human, robot’s appearance should be avoided so that can be also avoided any preconceived ideas about robots.

Telenoid, as a minimalistic human, was created following processes to remove as many unnecessary features as possible: (1) Choose the necessary features for communication from humans and discard the unrelated ones. (2) Choosen features are reconsidered to fit the requirements of the design of the Telenoid, deleting unnecessary features. (3) Essential features are obtained.

Unnecessary features were found from the researchers’ experiences in previous studies related to Geminoids. For example, Geminoid HI-1 can move its whole body: arms, legs, fingers and so on. On the other side, Geminoid F can move only its upper part of the body. But the two Geminoids are almost equally capable of conveying a specific human’s presence. This phenomenon indicates that body movements (except for facial ones) might be not so important for tele-

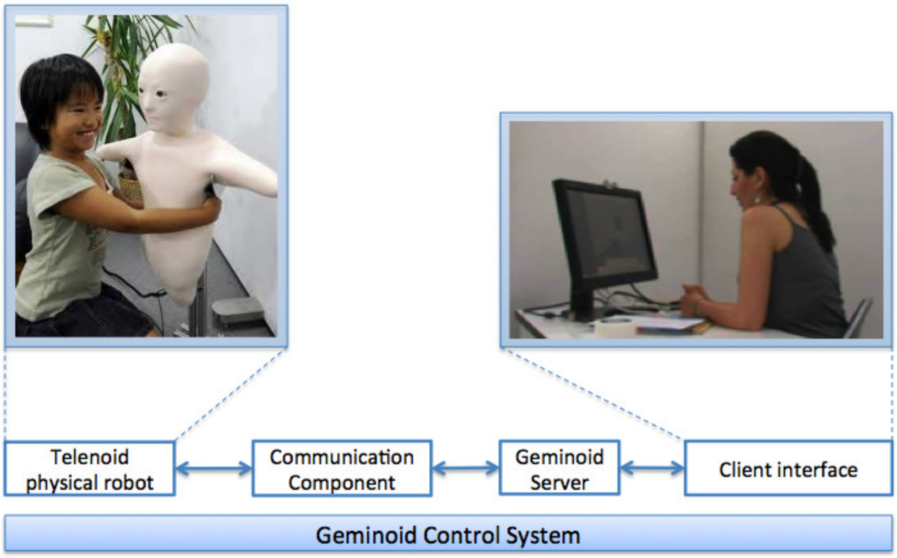

Figure 3.1: The control system and its components.

communication with Geminoids. Essential features that remain after this pruning process might be helpful to create an efficient tele-communication media that can be used by all types of people.

There are three advantages of the Telenoid compared with the Geminoids. The first one is that any kind of human can transfer its own presence. The second one is that the Telenoid allows people to make a physical communication. The last one is that the Telenoid can be easily moved anywhere. This means that the Telenoid can convey a specific human presence like the Geminoids, but can also be used by everyone because of its minimal design. Furthermore the size of the Telenoid facilitates the physical communication and can be used anywhere.

In this paper we describe two field tests in order to investigate whether the concept of the Telenoid can be accepted by ordinary people.

3.3 The Telenoid software interface

The software system that drives the Telenoid is based on the control system developed for the Geminoid HI-1 and F and is shown in figure 3.1.

In order to adapt the existing software system to the hardware of the robot, which is simplified with respect to the number and the physical specifications of the joints controlling the two Geminoids, it was necessary to develop the device dependent software part and an interface for the client.

The device dependent part is called "Communication Component" and has the role of translate the output motor commands coming from the geminoid server, whose range of possible values is 0 - 255, to the range of the physical motors embedded in the robot. Those ranges are:

- Axis 0 goes from -4500 to 4500
- Axis 1 goes from -4500 to 4500
- Axis 2 goes from -4900 to 4900
- Axis 3 goes from 0 to -4000
- Axis 4 goes from -8000 to 8000
- Axis 5 goes from 7000 to -7000
- Axis 6 goes from 9000 to -9000
- Axis 7 goes from 0 to -15000
- Axis 8 goes from 0 to 15000

The "Communication Component" for the Telenoid robot implements the following methods:

- *public int connect_robot()* This method is used to establish a connection with the robot. It returns 0 if the connection is successful.
- *public boolean disconnect_robot()* This method is used to disconnect the software layer from the robot. It returns "true" if the operation is successful.
- *public boolean motion_start()* This method enables the motion of the servos of the robot. Returns "true" if the operation was successful.
- *public boolean motion_stop()* This method puts the servos of the robot in "stop" state so that they cannot be activated by the software clients. Returns "true" if the operation was successful.
- *public boolean send_value_all(short[] val)* This method is called each 50 ms to set the position of each joint of the robot. It sends an array of values which will update the current state of each axis. It returns "true" if the operation was successful.

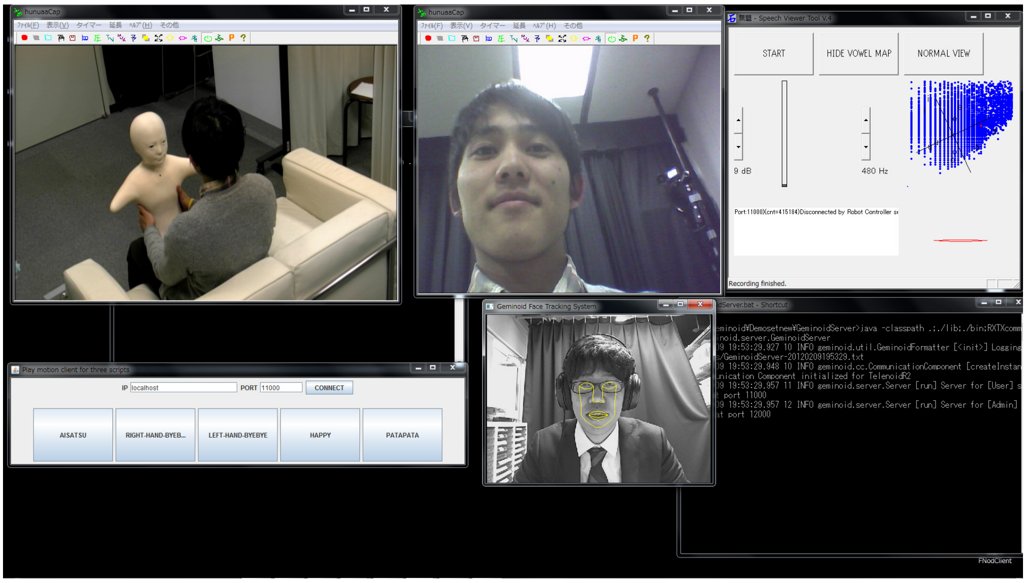

Figure 3.2: The software interface of the Telenoid.

A client interface has also been developed and it is shown in figure 3.2. Several windows are placed on the client screen.

On the top row the first window shows a third person view of the robot and his conversational partner. This kind of view is useful to the operator to gather spatial information about the robot and his conversational partner, such as their reciprocal distance, the position and orientation of the robot and a visual feedback of the body motions.

The second window is the output from the camera placed on the chest of the robot. This view is helpful because the camera points to the face of the interacting partner giving a clear understanding of his facial expressions.

In the third window there is the output of the lip-motion software module, which uses a method based on the rotation of the vowel space by using some specific features like the first and the second formants around the center vowel and mapping to the lip opening degrees Ishi et al. (2011).

In the bottom part of the screen there is a window which presents some preset behaviors for the robot, like show a bye bye motion with his left or right hand, or make a bow. The monitor used is a touch screen, so the operator can easily make his selection without move the focus from the conversation.

There is another window which shows the result of the face tracking software module, which is a commercial software called FaceAPI from Seeing Machines (http://www.seeingmachines.com). This part of the system tracks the head movements of the operator so that the robot can replicate them. The recognized face features are drawn with yellow lines.

Figure 3.3: Left: Aged person with Telenoid, Right: Demonstration at Design-Touch

In the last window, the debug information coming from the geminoid server are shown to the operator.

3.4 Demonstration at Shopping Mall

The first field test targeted ordinary people visiting the shopping mall. This field test was held as a part of an art event at shopping mall. We displayed the Telenoid for two days during this event and we conducted visitor interviews (fig: 3.3, right side).

3.4.1 Settings

At the beginning of the demonstration we provided basic information about how to use and communicate with the Telenoid (e.g. “This robot is a huggable and communicative medium. Please try to use it.”). After the explanation, we asked some applicants to sit on a sofa next to the Telenoid, then they started to talk with it. The duration of each conversation with the Telenoid was about 5 minutes. The operator was the only person who was accustomed to operate the Telenoid. The first field test targeted ordinary people interested in media art visiting the event. These visitors were highly interested in art and new technology. We asked 75 people to take an interview but some of them refused due to lack of time, so we collected a total of 56 interviews. These people were mostly in their 20’s (10’s: 6, 20’s: 30, 30’s: 12, 40’s: 4, 50’s: 2, unknown: 2). The interviewer took notes for recording the opinions of the visitors.

3.4.2 Interviews

We asked the visitors the following three questions. Interviews were conducted using natural conversation, so that visitors' opinions were hard to clearly divide into particular categories. Therefore the collected results were examined by three judges, and a visitor's opinion was classified only when the board voted unanimously, otherwise was classified as "neutral" (e.g. "strange and realistic" it's an answer to Q.1).

We show typical opinions for each question and describe brief tendencies for the opinions.

Q.1 How did you feel toward the Telenoid?

Typical opinions for Q.1 are as follows: (1) "At first glance, I felt strange but once I talked with it, I began to feel it was cute", (2) "Anyway I was just really scared", (3) "I felt attachment while I was talking with it".

The tendency of opinions for Q.1 shows that about half of the interviewed visitors felt positive (48.2%) and the other half felt negative (35.7%) or neutral (16.1%) toward the Telenoid. The typical negative opinion shows that the Telenoid's appearance is difficult to accept at first glance for ordinary people.

As a remarkable point 8 out of 11 visitors who answered "negative" for Q.1 mentioned that, after giving the Telenoid a hug, they felt positively toward the Telenoid. This result shows that the visitors who had negative impressions toward the Telenoid changed their mind by giving the Telenoid a hug.

Q.2 Was the Telenoid better than a telephone for talking with a distant person?

Typical opinions for Q.2 are as follows: (1) "I felt like I was in a space shared with the person operating the Telenoid", (2) "It might be easier to picture others using the Telenoid than using a telephone", (3) "When conversing with someone using the Telenoid, I can imagine the other person's emotion".

The tendency of opinions for Q.2 shows that 72.9% of interviewed visitors thought that the Telenoid was better than the telephone as a tele-communication media.

Q.3 Was it better to talk to a distant person using the Telenoid than to talk face to face?

Typical opinions for Q.3 are as follows: (1) "Direct communication is absolutely much better for me", (2) "I think I can speak my wife honestly by using the Telenoid", (3) "I can accept the Telenoid as a toy".

The tendency of opinions for Q.3 shows that 74.4% of interviewed visitors thought that face-to-face conversation is better than Telenoid.

3.4.3 Discussion

Concerning Q.1, about half of the interviewed visitors accepted the appearance of the Telenoid at first glance. At the same time the other half of the visitors, who had a negative impression of the Telenoid, changed their opinions after giving the Telenoid a hug. The Telenoid, as a minimal android, is a new concept. Nobody has had experience with the Telenoid. People tend to have negative impressions for new things. Therefore, it is not unexpected that almost half of the visitors felt strange using the Telenoid. However, hugging it decreased their negative impressions.

Next, the majority of interviewed visitors answered that the Telenoid is better than a telephone for talking with someone. On the other hand, a majority of interviewed visitors chose face-to-face conversation compared with the Telenoid. These results show that the Telenoid is currently not a replacement for face-to-face communication. The Telenoid has possibilities as a new tele-communication media. In the Q.3, one visitor answered “I can accept the Telenoid as a toy”. We classified the opinion into “negative” category, but we expected the opinion includes interesting knowledges. The visitor did not accept the Telenoid as a tele-communication medium, even though the visitor understood the concept of the Telenoid. It means that some people accept the Telenoid as only a conversational “agent”. In fact, three other visitors had similar opinions with it. This result shows that the Telenoid didn’t tread as a communication medium for whole kind of people.

The results of the field test might be biased by visitors’ familiarity with new technologies, because the field test was a part of an art event. However some of ordinary people, whose purpose were not the art event, were mixed with visitors because the art event was took place inside of a common shopping mall. This might support a certain generality of the results.

As mentioned above, people tend to avoid new unknown things. If the Telenoid becomes popular in the future, we hope that ordinary people might take advantage of using the Telenoid when they want to feel the presence of a distant person.

3.5 Telenoid with Aged Persons

The second field test involved elderly persons using the Telenoid (fig: 3.3, left side). This second experiment provided us with further material for discussion.

3.5.1 Settings

This test was held as a part of a tour introducing our laboratory. The aged persons went to several locations, including one where the Telenoid demonstration was held. This exposure might have allowed some of them to become accustomed to the environment. We interviewed 47 aged persons. They were in their 70's, 80's and 90's (70's: 6, 80's: 26, 90's: 9, unknown: 6). The interviewer used videos for recording the opinions of visitors. They are receiving a service from a "Day Care Center" at their own home and they were not diagnosed with dementia.

First, we provided basic information regarding the Telenoid (e.g. "This robot is a communication medium. Please try to use it."). After the explanation, some applicants were asked to sit on a sofa placed next to the Telenoid, then they started talking using the Telenoid. The duration of conversation with the Telenoid was about 5 minutes. The operator was a female employee from the Day Care Center. The staff and aged persons were acquainted.

3.5.2 Interviews

We asked the following 4 questions to the aged persons. Interviews were conducted while engaging in natural conversation, so the opinions obtained are hard to clearly divide into particular categories. The collected results were examined by three judges, and a visitor's opinion was classified only when the board voted unanimously, otherwise was classified as "neutral" (e.g. "it is bit heavy and looks like its mother" it's an answer to Q.2).

Additionally, answers from the visitors were sometimes not consistent because of their advanced age. Therefore some inconsistent answers, such as "I like moni-chan" (it's an answer to Q.2, we could not catch the meaning of "mochi-chan"), were removed from the results because it's difficult to classify into categories.

Q.1 could only be answered as "staff" or other. Thus the description regarding typical opinions for Q.1 is skipped.

Q.1 Whom were you talking with?

The aged persons and the staff have known each other. However the result for Q.1 showed that about 47% of the elderly did not realize whom they were talking with through the Telenoid even though the experimenter had described the Telenoid before the interaction.

Q.2 How did you feel toward the Telenoid?

Typical opinions for Q.2 are as follows: (1) "Very cute. It looks like my grandchild", (2) "It is very soft and nice to touch", (3) "It is very far from actual humans and the tactile sensation is like rubber".

The tendency of opinions for Q.2 shows that 88.8% of aged persons felt positively toward the Telenoid. As a remarkable point, all aged persons gave the Telenoid a hug without any specific instructions (e.g. "This robot is a huggable media. Please try hugging it."). When an experimenter handed the Telenoid to an aged person, they gave the Telenoid a hug immediately, and they started talking to the Telenoid with gentle stroking. They seemed happy to interact with the Telenoid.

Q.3 Is the Telenoid good to talk with a distant person compared with the telephone?

Typical opinions for Q.3 are as follows: (1) "I think I can feel the actual person's presence with the Telenoid", (2) "I am not used to using the Telenoid; the telephone is better for me", (3) "The Telenoid is better because it is very cute".

The tendency of opinions for Q.3 shows that 66.6% of aged persons had a positive opinion of the Telenoid. For example, the answer (1) and (3) were classified into "positive" group, and the answer (2) was classified into "negative" group.

Q.4 Is the Telenoid good to talk with distant person compared with face to face conversation?

Typical opinions for Q.4 are as follows: (1) "Direct conversation is better. Because humans are alive", (2) "Face to face is good for me", (3) "I like to talk with my grandchild face to face".

The tendency of opinions for Q.4 shows that about 26% of aged persons had a positive opinion of the Telenoid. For example, three typical opinions were classified into "negative" group.

3.5.3 Discussion

Concerning the result of Q.1, almost half of the aged persons did not realize who was operating the Telenoid. This suggests that maybe the concept of tele-operation was slightly difficult to understand for them. However, as the results of Q.2, the aged person had a positive impression of the Telenoid from the very start. It means that the Telenoid could not be accepted as a tele-communication medium to some aged persons, but it has potentials to accept as the other medium, which is for talking.

The results of Q.3 and Q.4 indicate a similar tendency to that seen in the last test. For aged persons as well, it is hard to imagine the Telenoid as being a replacement for humans (face to face communication). But as a tele-communication

Figure 3.4: Photos of the experiment at DINFO.

media, Telenoid is acceptable for them.

The most different point compared with younger people's reactions was the manner of conversation. For example, some aged persons did not pay very close attention to what the Telenoid said, and they instead talked about their own experiences. It might seem that Telenoid did not work efficiently for the elderly visitors. However they hugged the Telenoid and showed big smiles when talking with the Telenoid. Although the Telenoid is basically intended for tele-communication, some of the aged persons treated the Telenoid as just a huggable and communicative "agent". We think that this could also be a proper way to use the Telenoid. The most important goal for this Telenoid research is to discover not only unknown possibilities for the Telenoid but also for androids as a whole. The knowledge which we obtained from this field test with aged persons might be useful for android studies in the future.

3.6 Demonstration at DINFO

In September 2011 a set of experiment has been performed at DINFO department of the Universita' degli Studi di Palermo (shown in figure 3.4).

3.6.1 Theoretical Remarks and Modeling

A principled approach to human-robot interaction may be assumed to comply with the natural conditions of agents overt perceptual and social behavior. Arbib and Fellous (2004) argue that human-robot interaction settings, and machine simulations of cognitive abilities provide with a novel and meaningful two-way research into behavior organization,social coordination and communication of both animal and artificial agents. Indeed, we take humanoid robots to give test beds, and the design of human robot interaction to prove an effectual strategy to study those perceptual and social abilities of agents copying with environment, which can also be justifiably presumed to have robots appear to sense and behave like humans, and accordingly to give a human-like character to human-artifact interaction.

Much relevant literature appeared on the features of natural character of agents interaction. Since Argyle (1994), Argyle and Cook (1976) research was devoted to specify the different functions of gaze, and proxemic indicators that contribute to the organization and the dynamics of the interpersonal space that subserves social cognition and behavior (the so-called "equilibrium theory" Argyle and Dean (1965)). On these grounds, Blascovich et al. (2002) discuss the theoretical framework laid bare for these features and its application to shared immersive virtual environments. Torres et al. (1997) apply the empirical analysis of gaze behavior in a dyadic-conversation paradigm to show a meaningful relationship between gaze and information retrieval of discourse content, in order to devise a model for an algorithm that retrieves coupling of gaze directions and meaningful parts of propositional contents along with utterance attributes in the communicative humanoid agent proposed by **?**. Vertegaal et al. (2001) argue that evidence of gaze function in coordinative behavior comes from research on gaze directional clues as reliable non-verbal predictors of conversations in multi-agent, multi-user environments. Mutlu et al. (2006) study the extent at which gaze contact frequency among a storytelling robot and its human listeners is correlated with story understanding and recall, and approval ratings in the evaluation of robot performance. They argue that results highlight meaningful commonalities between human-human and human-robot communication.

Jackson and Decety (2004), and Pietroni et al. (2008) drew attention to other features of overt behavior that play an informative role in understanding other human agents' purposive and intentional behavior, that is motor behavior and displayed emotions particularly in contexts of coordinative or competitive behavior, when such overt features as gaze behavior, eye contact, emotions display give agents proxies of behavior perceptual proxies to detect what another agent is looking at and is directing her attention to, and why. Indeed, that cognitive integration of perceptual features seem to ascription and recognition of intentions to obtain.

A common framework for all those research lines was already set by Wicker et al. (2003) who asked subjects to attribute hostile or friendly intentions to videotaped actors who directed attention towards or away from the subjects. The aim was to identify a brain system associated with perception and attribution of emotion displayed in the eye region in the specific context of direct gaze. They found that this information processing turns out to be different when emotions and intentions of other agents are perceived as directed at the subjects or elsewhere. Different brain areas recruitment and different level of activations are reported to occur when individuals experience and judge the emotional nature of a gaze compared to a neutral gaze, as well as when subjects experience and judge an emotional direct gaze versus an emotional averted gaze. Neurophysiological research (as early as Anderson et al. (1991)) achieved consensus about findings that attest the crucial link among perception of features, be they emotional or gaze qualities of interacting agents behavior, and representation of emotional and social significance of salient stimuli by tracking the coordinated activation in specialized cortical and subcortical systems (Adolphs (2002); Phan et al. (2002); Phillips et al. (2003)). We reasoned that the emerging picture of natural interaction condition requires a general description of the environment where agents cognitive and social interaction with their surroundings that can be carved up at the level at which environment look somehow like to agents. Since Koffka (1955), Köhler (1992), Heider (1982), and Lewin et al. (1936), cognition and behavior is proposed to be analyzed at the scale of what agents themselves take as meaningful units and accordingly their environment can be decomposed in what they see as directly or indirectly accessible objects sense properties, affordances, scaffolds and proxies of other agents intentions and behaviors. Analytical treatment of the behavioral environment as it looks like from an agent's standpoint allows to recover its qualitative structure that support cognition, agency and interaction with other agents to cope efficiently by trading off her own decision making and action course selection against resources, dangers, opportunities, and observed other's behavior. As Chrisley (2009) pointed out, there is no hindrance to the definition of a "synthetic phenomenology" devoted to the research of perceptual qualities that carry out cognitive functions even for the field of IA and robotics.

Indeed that approach may deliver a theoretical gain. Agents behavior is to be explained as organized and regulated by the cognitive frames of reference that build up their environment, in that perception of objects, events, and other agents is connected with action, goals attainment, and competitive or cooperative coordination with other agents. Conditions for natural behavior are to be specified at a meso-scale level where they are accessible as they look like to agents.

3.6.2 Interaction setting and research design

From this theoretical framework and the relevant literature, we derived some features of overt behavior that qualify as parameters for efficacious interaction given that they can work as perceptual and cognitive shared blocks of the environment where agents interact. We profited from Telenoid, a humanoid robot endowed with some of perceptual and motor features of overt behavior tuned to speech and head movement of a human agent through a tele-operated system, to study such perceptually accessible features as meaningful clues for social interaction. A model of interaction was then set:

- (1) to understand how some perceptual features work, such as distance and relative positions of agents, face regions spotted as highly informative about emotion or intention reading, the degree at which the space where the interaction obtain appears to be a shared environment;

- (2) to recover which perceptual features of overt behavior among head movements, gaze and eye contact search, presence or absence of lips movement are held as salient by human agents to ascribe a meaningful conduct to the robot;

- (3) to assess the degree of believability of such an interaction along dimensions that can be reasonably taken as meaningful indicators of social interaction, both in free and task directed conditions. Hence, the interaction setting and the research methodology were modeled as followed.

3.6.3 Participants

Industrial Design Course students of the Faculty of Architecture (University of Palermo) were recruited who did not have prior interaction experience with humanoid robots, though it was not excluded that they possessed informations or informal notions of IA and robotics. There was no selection process. They were only informed of the possibility of taking part in a robot research, and those who freely declared to have an interest whatsoever in joining it were selected. All participants were given a brief explanation of Telenoid, of the interaction setting structure that required a two stage interaction with the robot and to fill up a questionnaire. Each participant was asked to choose a partner for the interaction and then to decide who will be interacting and who will be tele-operating the robot. A small number of couples of participants was allowed to switch their member role between the two stages of interaction. Each couple was introduced to the setting and the control box by one of the researcher who was appointed also to tell participants when each interaction stage was deemed to be over. All interaction were videotaped.

3.6.4 Design

A two stage interaction with Telenoid was prepared. A first free interaction stage, meant to allow subjects to adapt to interact with the humanoid robot and, for those chosen to be tele-operating, to acquire as early as possible the skills for operating the robot through the control box. The subject that chose to teleoperate the robot entered a separate room where the control box was located in order to have him not visible to the other subject that interacted directly with to robot.

A second interaction stage was instead task driven. Participants were allowed to choose an interactions scenario among a proposed range that spanned booking a hotel reservation, making a phone call to a mobile company to get contract or services information, to matriculate or to enter his/her name or one of his/her fellows ones for a course examination by talking directly with the robot. Those who were to tele-operate the robot were asked to use all the knowledge acquired in this standardized context of interaction to act as formal as possible. This second stage was stopped as soon as the goal was attained. Before the first stage of interaction, each subject was asked to fill up a first part of the questionnaire whose questions range over general information about his/her own interest and hobbies, his/her interest in such fields as technology and robotics, his familiarity or knowledge of such fields, and the implicit degree of acceptability of robotic artifacts. After the first stage of interaction, each subject was given a free test that was assumed to serve as a distractor in order to avoid that expectations arisen after the adaptation interaction could distort the direct experience in the task driven second stage. The free test consisted in a random presentation of humanoid and not-humanoid robots pictures to whom each subject should systematically couple an emotion name from a fixed set provided in a paper list. After the second task guided stage of interaction, subjects were requested to fill up the second part of the questionnaire whose questions were about those very constructs built to recover information about the salient perceptual and social dimensions of the interaction. A few days later the interaction setting, a third questionnaire was administered only to subjects who tele-operated the robot.

3.6.5 Methodology: constructs and item analysis

The questionnaire was meant to cover two main constructs that according to theoretical assumption could recover some perceptual and social aspects of natural conditions of agents interaction, which were also hypothesized to rule the human-humanoid robot interaction. We reasoned that those aspects mirrored some salient ordinary cognitive abilities, which agents could specify in such cases to improve the efficacy of interaction. The first construct is intended to cover perceptual features of overt interactive behavior. It is represented by items that add up to three different but convergent aspects: (1) the apparent distance of agents, and their sense of being sharing a common environment; (2) the perceptual attention to those parts of robotic device that are perceived as more likely displaying meanings and intentions; (3) the reliability of robot's observable behavior given the degree of consistency among head and arms movements, gaze, utterance synchronization. Groups (1) and (2) are provided also with items asking subjects whether they needed to change some perceptual parameter in order to improve the interaction. Questions about the perception of distance were designed in order to cover in ordinary and informal way findings about the multiple functions that space regions have when endowed with perceptual and motor interpretation to carry out or detecting meaningful action (neurobiological evidence about the motor-cognition integration systems that decompose interaction space in multiple phenomenal maps is summed in Gallese (2005)). For theoretical reasons, the perceptual awareness of sharing a common environment with the robot is assumed to be of momentous importance for the interaction subjects to ascribe intentions and actions to the robot itself and not to it only as an apparent proxy of the tele-operating subject. This aspect of interaction can prove to be the perceptual link with the second construct: believability. The concept is defined in Dautenhahn (1998), and Poel et al. (2009) operationalize it designing a construct whose aspects are represented by item grouped according to the indicators of personality, emotion, responsiveness, and self motivation. We chose to construe believability along the following dimensions: (4) valence, whose items cover the apparent robot capability to act due to internal or external (other agents, environment) causes; (5) motivation, whose items cover the apparent robot capability to display interest in other agents requests and goals; (6) value, whose items cover the coordination of robot's behavior with human agents; (7) communication, whose items cover the coupling between the robot overt behavior and the intentions ascribed to it. These questionnaire parts were administered after the second task driven stage of interaction. Hence they are meant to represent the perceptual and social dimension of the more complex interaction given that the impressions that struck subjects eventually in the first free interaction had only an adapting subserving function for an efficacious interaction to obtain. These two parts were given together with a third part of the questionnaire that was about the assessment of the overall interaction with the robot. The third questionnaire, which was administered a few days later only to those subjects who tele-operated the robot, was made up of questions that spanned the assessment of the technical design of the control box, and its usability as regards the transmission of subjects' own head movement or the delay between subjects' utterances with respect to their perception by users interacting with the robots. Some eventual suggestions to improve usability were also included in order to make subjects operating intentions clear to those who interacted directly with the robot. Furthermore the technical features of the device were often traded off with perceptual and cognitive constraints that were presumed to constrain subjects task to drive the robot in such a way to interact effectively with other subjects and at the same time to have it appear as autonomous as possible. Items were formulated in the form of a single forced or multiple closed choice set, and as statements for which subjects were to rate agreement on a five-point Likert scale. We assumed that the perceptual and social dimensions can be reasonable represented as continua on a multiple items attitude scale. Standard item analysis has been performed on the codified data. Split-half Spearmann-Brown coefficient and

Cronbach Alpha were meant to test the reliability of the scales and the correlation among the multiple items of each single construct. We found significantly high values of reliability as regards the items correlation and the internal consistency of the scales measuring the perceptual and social dimensions. Given the substantive assumption the led to the design of our questionnaire and the high number of items per scale, high reliability value may depend on their number and particular choice which fit the theoretical structure we wanted the interaction model to have. Further analysis is needed to decompose the meaning of our results, but for the time being the extent at which each construct covers the multiple aspects of the assumed dimension is reasonably to be taken as achieved.

3.6.6 Expected results and future research

The present research aims at a descriptive analysis of the main perceptual and social features of natural conditions of agent interaction, which can be specified by agent in human-humanoid robot interaction. We maintain that such a descriptive research can highlight dimensions that contribute to meaningful and natural-like interaction with humanoid robots. The main upshot of our research is the definition of two dimensions that can be taken to underlie human experience with interacting artifacts in ordinary and goal driven contexts. Those dimensions are specified as perceptual and social features according to theoretical assumptions on the level at which descriptive units of behavior must be detected. They can be construed as multiple items that can be reliable indicators to retrieve some of those perceptual and social behavior skills agent realize when faced with interaction contexts. As regards the current literature, we focus mainly to perceptual features of agents and multiple spaces for interaction behavior, which can be perceptually specified in motor and action terms to an effectual agents interaction to obtain. And for as social dimension is concerned, we construed believability as more linked to some features of observable behavior, which can prove momentous for agents disposition to social coordination, than as they are in the current literature, at least at the light of our current knowledge.

Given the data analysis performed, we may claim only to have individuated two interaction dimension that can serve for either assess the perceptual and observable behavior conditions an humanoid agent is justifiably taken to comply with, or to increase the natural-looking-like of interaction behavior in human humanoid interaction. In order to specify their single meaning, that is to quantify over, the degree at which they represent a specification by of unified conditions agents implicitly comply to, the extend at which these dimensions mutually reinforce each other and give a joint contribute to a successful and effectual interaction to obtain further analysis and research are required. By administrating or questionnaire to subject after well defined experimental conditions, an analysis of significant variance correlation among dimensions in ordinary and goal guided context of interaction may be performed. Furthermore, by coupling interaction setting conditions with brain imaging techniques, or ERP registration our descriptive analysis is likely to upgrade to a promising explanatory level.

and their integration in cognitive modules of apparent behavior for social coordination and believability. Given the limitations of our interaction setting design and methodology, future research can be devoted to design experimental conditions to be coupled with question techniques to upgrade to a promising explanatory level.

3.7 Conclusions and Future Works

This thesis proposed a hypothesis for the improvement of the human likeness of an android by establishing, breaking, and recovering the eye contact with a human interlocutor, as well as studying the essence of the human presence using a telepresence media robot like the Telenoid.

Those social signals can be used in both face-to-face interactions and long range interactions, and they can reinforce our expectation of androids as a responsive agent.

Previous studies were focused on the design and development of non conventional user interfaces like the BCI Brain-Computer Interface for the control of a robot in an experiment with the San Camillo Hospital in Padova and the DINFO Department of the University of Palermo. This work can be further improved and applied to the Geminoid robots.

Also, other works like the one in **?** focused on the study of the emotional robotics, which aims to give a robot internal emotional states which can enrich the interaction with users.

3.8 Contributions

The main contributions of the work presented in this dissertation are:

- bla bla bla ....
- bla bla bla ....

3.9 Personal publications

The work in this thesis has been published in several referred conference proceedings:

- K.OGAWA, S.NISHIO, K.KODA, G.BALISTRERI, T.WATANABE, and H.ISHIGURO (2010). Exploring the Natural Reaction of Young and Aged Person with Telenoid in a Real World. Journal of Advanced Computational Intelligence & Intelligent Informatics (JACIII)

- S.M. ANZALONE, P. BALISTRERI, SORBELLO R., A. CHELLA (2010). An emotional robotic partner for entertainment purposes. INTERNATIONAL JOURNAL OF COMPUTATIONAL LINGUISTICS RESEARCH, vol. 1; p. 94-104, ISSN: 0976-416X.

- A. CHELLA, R. SORBELLO, G. PILATO, G. VASSALLO, G. BALISTRERI, and M. GIARDINA (2011) An Architecture with a Mobile Phone Interface for the Interaction of a Human with a Humanoid Robot Expressing Emotions and Personality. Lecture Notes in Artificial Intelligence (LNAI), n. 6934, AI*IA 2011: Artificial Intelligence Around Man and Beyond,XIIth International Conference of the Italian Association for Artificial Intelligence, Palermo, Italy, September 15-17, 2011. Proceedings, Springer-Verlag Berlin Heidelberg, pp.117-126, ISBN: 978-3-642-23953-3.

- A. CHELLA, R. SORBELLO, G. PILATO, G. BALISTRERI, S.M. ANZALONE (2011). An Innovative Mobile Phone Based System for Humanoid Robot Expressing Emotions and Personality . In: Frontiers in Artificial Intelligence and Applications (FAIA) Biologically Inspired Cognitive Architectures. 2011 Proceedings of the Second Annual Meeting of the BICA Society A.V. Samsonovich et al. (Eds.). vol. 233, p. 57-63, IOS PRESS, ISBN/ISSN: 978-1-60750-958-5.

- G. BALISTRERI, S. NISHIO, R. SORBELLO, A. CHELLA AND H. ISHIGURO. Natural Human Robot Meta-Communication Through the Integration of Android's Sensors with Environment Embedded Sensors. In: Frontiers in Artificial Intelligence and Applications (FAIA) Biologically Inspired Cognitive Architectures. 2011 Proceedings of the Second Annual Meeting of the BICA Society A.V. Samsonovich et al. (Eds.). vol. 233, p. 26-38, IOS PRESS, ISBN/ISSN: 978-1-60750-958-5.

- G. BALISTRERI, S. NISHIO, R. SORBELLO, AND H. ISHIGURO (2011). Integrating Built-in Sensors of an Android with Sensors Embedded in the Environment for Studying a More Natural Human-Robot Interaction. Lecture Notes in Artificial Intelligence (LNAI), n. 6934, AI*IA 2011: Artificial Intelligence Around Man and Beyond,XIIth International Conference of the Italian Association for Artificial Intelligence, Palermo, Italy, September 15-17, 2011. Proceedings, Springer-Verlag Berlin Heidelberg, pp.432-437, ISBN: 978-3-642-23953-3.

Bibliography

R. Adolphs. Neural systems for recognizing emotion. *Current opinion in neurobiology*, 12(2):169–177, 2002.

SW Anderson, A. Bechara, H. Damasio, D. Tranel, and AR Damasio. Adolphs, r.(1999). social cognition and the human brain. trends in cognitive science, 3, 469-479. allen, c.(2004). transitive inference in animals: Reasoning or conditioned associations? in sh and m. nudds (eds.), rational animals? oxford. *Behavioral and Brain Sciences*, 14:471–517, 1991.

M.A. Arbib and J.M. Fellous. Emotions: from brain to robot. *Trends in cognitive sciences*, 8(12):554–561, 2004.

M. Argyle. *The psychology of interpersonal behaviour.* ePenguin, 1994.

M. Argyle and M. Cook. Gaze and mutual gaze. 1976.

M. Argyle and J. Dean. Eye-contact, distance and affiliation. *Sociometry*, pages 289–304, 1965.

C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita. Is the uncanny valley an uncanny cliff? In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 368–373. IEEE, 2007.

C. Becker-Asano, K. Ogawa, S. Nishio, and H. Ishiguro. Exploring the uncanny valley with geminoid hi-1 in a real-world application. In *IADIS Intl. Conf. Interfaces and Human Computer Interaction*, pages 121–128, 2010.

S.J. Blakemore and U. Frith. How does the brain deal with the social world? *Neuroreport*, 15(1):119, 2004.

J. Blascovich, J. Loomis, A.C. Beall, K.R. Swinth, C.L. Hoyt, and J.N. Bailenson. Target article: Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, 13(2):103–124, 2002.

R.A. Brooks, C. Breazeal, M. Marjanović, B. Scassellati, and M.M. Williamson. The cog project: Building a humanoid robot. In *Computation for metaphors, analogy, and agents*, pages 52–87. Springer-Verlag, 1999.

A. Bruce, I. Nourbakhsh, and R. Simmons. The role of expressiveness and attention in human-robot interaction. In *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, volume 4, pages 4138–4142. IEEE, 2002.

T. Chikaraishi, T. Minato, and H. Ishiguro. Development of an android system integrated with sensor networks. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 326–333. IEEE, 2008.

R. Chrisley. Synthetic phenomenology. *International Journal of Machine Consciousness*, 1(1):53–70, 2009.

K. Dautenhahn. The art of designing socially intelligent agents: Science, fiction, and the human in the loop. *Applied Artificial Intelligence, 12*, 7(8):573–617, 1998.

V. Gallese. Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, 4(1):23–48, 2005.

F. Heider. *The psychology of interpersonal relations.* Lawrence Erlbaum, 1982.

M. Imai, T. Kanda, T. Ono, H. Ishiguro, and K. Mase. Robot mediated round table: Analysis of the effect of robot's gaze. In *Robot and Human Interactive Communication, 2002. Proceedings. 11th IEEE International Workshop on*, pages 411–416. IEEE, 2002.

C.T. Ishi, C. Liu, H. Ishiguro, and N. Hagita. Speech-driven lip motion generation for tele-operated humanoid robots. In *Auditory-Visual Speech Processing 2011*, 2011.

H. Ishiguro. Distributed vision system: A perceptual information infrastructure for robot navigation. In *International Joint Conference on Artificial Intelligence*, volume 15, pages 36–43. Citeseer, 1997.

H. Ishiguro. Android science: conscious and subconscious recognition. *Connection Science*, 18(4):319–332, 2006.

H. Ishiguro. Android science–toward a new cross-interdisciplinary framework. *Robotics Research*, 28:118–127, 2007.

P.L. Jackson and J. Decety. Motor cognition: a new paradigm to study self-other interactions. *Current Opinion in Neurobiology*, 14(2):259–263, 2004.

T. Kanda, H. Ishiguro, T. Ono, M. Imai, and R. Nakatsu. Development and evaluation of an interactive humanoid robot robovie. In *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, volume 2, pages 1848–1855. IEEE, 2002.

T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro. Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19(1):61–84, 2004.

K. Koffka. *Principles of Gestalt psychology.* New York, 1955.

W. Köhler. *Gestalt psychology: An introduction to new concepts in modern psychology.* Liveright Publishing Corporation, 1992.

K. Lewin, F.T. Heider, and G.M. Heider. Principles of topological psychology. 1936.

K.F. MacDorman and H. Ishiguro. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3):297–337, 2006.

K.F. MacDorman, T. Minato, M. Shimada, S. Itakura, S. Cowley, and H. Ishiguro. Assessing human likeness by eye contact in an android testbed. In *Proceedings of the XXVII annual meeting of the cognitive science society*, pages 21–23, 2005.

Y. Matsumoto, T. Wada, S. Nishio, T. Miyashita, and N. Hagita. Scalable and robust multi-people head tracking by combining distributed multiple sensors. *Intelligent Service Robotics*, 3(1):29–36, 2010.

Y. Matsusaka, T. Tojo, S. Kubota, K. Furukawa, D. Tamiya, K. Hayata, Y. Nakano, and T. Kobayashi. Multi-person conversation via multi-modal interface-a robot who communicate with multi-user. In *Sixth European Conference on Speech Communication and Technology*, 1999.

T. Minato, M. Shimada, H. Ishiguro, and S. Itakura. Development of an android robot for studying human-robot interaction. *Innovations in applied artificial intelligence*, pages 424–434, 2004.

M. Mori. The uncanny valley. *Energy*, 7(4):33–35, 1970.

T. Mori, H. Noguchi, and T. Sato. Daily life experience reservoir and epitomization with sensing room. In *Proceedings of Workshop on Network Robot Systems: Toward Intelligent Robotic Systems Integrated with Environments*, 2005.

H. Morishita, K. Watanabe, T. Kuroiwa, T. Mori, and T. Sato. Development of robotic kitchen counter: kitchen counter equipped with sensors and actuator for action-adapted and personally fit assistance. In *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 2, pages 1839 – 1844 vol.2, oct. 2003. doi: 10.1109/IROS.2003.1248911.

B. Mutlu, J. Forlizzi, and J. Hodgins. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*, pages 518–523. IEEE, 2006.

S. NISHIO and H. ISHIGURO. Attitude change induced by different appearances of interaction agents. *Int. J. Mach. Conscious.*, 3(01):115–126, 2011.

S. Nishio, H. Ishiguro, and N. Hagita. Geminoid: Teleoperated android of an existing person. *Humanoid robots-new developments. I-Tech*, 2007.

K.L. Phan, T. Wager, S.F. Taylor, and I. Liberzon. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in pet and fmri. *Neuroimage*, 16(2):331–348, 2002.

M.L. Phillips, W.C. Drevets, S.L. Rauch, and R. Lane. Neurobiology of emotion perception i: The neural basis of normal emotion perception. *Biological psychiatry*, 54(5):504–514, 2003.

D. Pietroni, G.A. Van Kleef, C.K.W. De Dreu, and S. Pagliaro. Emotions as strategic information: Effects of other's emotional expressions on fixed-pie perception, demands, and integrative behavior in negotiation. *Journal of Experimental Social Psychology*, 44(6):1444–1454, 2008.

M. Poel, D. Heylen, A. Nijholt, M. Meulemans, and A. Van Breemen. Gaze behaviour, believability, likability and the icat. *AI & Society*, 24(1):61–73, 2009.

F.H. Previc and J.M. SHANNON. Vertical eye movements during mental tasks: A re-examination and hypothesis. *Perceptual and motor skills*, 84(3):835–847, 1997.

D. Sekiguchi, M. Inami, and S. Tachi. Robotphone: Rui for interpersonal communication. In *CHI'01 extended abstracts on Human factors in computing systems*, pages 277–278. ACM, 2001.

M. Shimada, T. Minato, S. Itakura, and H. Ishiguro. Evaluation of android using unconscious recognition. In *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*, pages 157–162. IEEE, 2006.

E. Takano, T. Chikaraishi, Y. Matsumoto, Y. Nakamura, H. Ishiguro, and K. Sugamoto. Psychological effects on interpersonal communication by bystander android using motions based on human-like needs. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 3721–3726. IEEE, 2009.

O. Torres, J. Cassell, and S. Prevost. Modeling gaze behavior as a function of discourse structure. In *First International Workshop on Human-Computer Conversation*. Citeseer, 1997.

R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 301–308. ACM, 2001.

B. Wicker, D.I. Perrett, S. Baron-Cohen, and J. Decety. Being the target of another's emotion: a pet study. *Neuropsychologia*, 41(2):139–146, 2003.