Development of culturally-adaptive non-verbal communication capabilities for humanoid robots

ヒューマノイドにおける文化規範型非言語的コミュニケーション機能の開発

February 2014

Gabriele TROVATO

トロヴァト　ガブリエレ

Development of culturally-adaptive non-verbal communication capabilities for humanoid robots

ヒューマノイドにおける文化規範型非言語的コミュニケーション機能の開発

February 2014

Waseda University

Graduate School of Advanced Science and Engineering

Department of Integrative Bioscience and Biomedical Engineering

Research on Bio-robotics

Gabriele TROVATO

トロヴァト　ガブリエレ

Acknowledgements

I always imagined a Ph.D. course as a mountain that has to be climbed. Starting work in a new field, in a foreign country, with the requisite of 3 journal papers to be published in 3 years, sounded indeed like a tall mountain. In such a journey, it is difficult to understand where you are heading, since research means doing something new, thus following a path that has not been traced yet. However, once you are on the top, the view changes and you have the clear idea of what you did and where to go next.

I would like to thank all the people who supported me in this journey and made this achievement possible, beginning with Prof. Takanishi. His laboratory is like a big family: special thanks go to Prof. Zecca; Dr. Ishii; Dr. Hashimoto and the Biped group; the secretary Mrs. Ohta; the Italian group, especially Prof. Sessa, Dr. Bartolomeo, and Dr. Jamone.

In Waseda, I would also like to thank Prof. Kuramochi as well as the co-referees Prof. Umezu, Prof. Fujie, Prof. Hashimoto and Prof. Kimura.

Regarding the overseas universities visited in these years, the list is quite long, beginning with Prof. Asfour and all the members of the Humanoids and Intelligence Systems Lab in Karlsruhe Institute of Technology, in Germany. In the Netherlands, Prof. Ham and Ms. Liu in Eindhoven University of Technology, and Dr. Chrupala in Tilburg University. Then all the staff in E-JUST and in JICA, in Alexandria, Egypt. Finally, Prof. Waibel and the Interactive Systems Lab members in Carnegie Mellon University, Pittsburgh, USA.

These research trips were funded by Global COE Program "Global Robot Academia" from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and by the "Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation" program from the Japanese Society for the Promotion of Science. My extreme gratitude goes to Prof. Takanishi and Dr. Ishii for letting me join the program. My stay in Japan was funded by the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT).

I would like to thank also all the other people who have been close to me in these last years, including my family, and anyone else whom I may have forgotten.

Abstract

In the near future humanoid robots are expected to play a major role in the society, interacting and communicating with humans to help them in their work and daily life. Typical examples are serving as assistive robots for the elderly, or serving as companion robots. Assistive and personal robots are a class of devices whose design should be especially driven by considerations about human acceptance. Several studies have shown that the cultural background affects the attribution of some form of personality to the robots. The idea that acceptance of robots depends on the culture is a very sensitive issue: Asimov was the first to introduce the Frankenstein complex, which describes the anxiety that people feel towards robots. According to the traditional view in literature, such anxiety is in part caused by popular fictional stories in which robots have negative connotations (e.g. Frankenstein, Terminator), at least in Western countries. This complex seems to be absent in Japan, where robots are viewed more like helpers or heroes. As for the Middle East, few researches have been done so far in the field of human-robot interaction. However, in order to improve acceptance of robots and their penetration into this culture, the implications of religious beliefs, especially in the case of humanoid robots, should be considered. As stated by Thomas and Rogers, understanding cultural norms of a country is important not only for robotics, but it necessary for ensuring technology acceptance. Therefore, in human robot interaction it is important that robots adapt to the different ways humans use to communicate in different cultures.

Communication between two humans is achieved through the simultaneous use of both verbal and nonverbal communication. Robots too should be able to use these two channels and adapt them to the specific culture of the human partner.

Nonverbal behaviour is particularly important because it gives a tone to the whole communication, and sometimes can override the verbal part. As humans, we use different types of nonverbal cues, such as kinesics, proxemics, haptics, and paralanguage. For humanoid robots, kinesics is very important, as their human-like shape allows them to perform body and facial expressions during communication.

However, nowadays humanoid robots have limited capabilities in terms of non-verbal communication.

In the case of the face, while a few examples of robots that can already perform a certain number of facial expressions, the number of expressions is usually limited and the patterns are pre-defined. There is a need to go beyond the traditional approach of performing only the most basic expressions (fear, anger, disgust, happiness, sadness, and surprise), and rather use a parametrical approach that would make the robot able to generate facial expressions flexibly for a specific culture. This concept can be extended to expressions that are not strictly defined as emotions, but are rather classified as communication acts (such as incomprehension or rebuke).

As for body language, very few studies have investigated how robots speaking a language and using gestures that belong to a certain culture are perceived by humans with different cultural backgrounds. As a result, robots gestures are generally defined without taking into account the culture of the human partner.

The purpose of this thesis is to study how robots design and behaviour can be re-thought to comply different cultures and be perceived by humans as more familiar. This will ultimately lead to improve robots acceptance and reduce discomfort experienced during interaction. This is fundamental for future deployment of robots into the society, in different parts of the world. In order to fulfil this purpose, 4 concrete objectives are then set:

a. develop a flexible facial expression generation system that can be applied to the 24 degrees of freedom head of the humanoid robot KOBIAN-R and makes possible the generation of facial expressions from composite emotions and communication acts;

b. measure the possible gap in recognition of facial expressions of the robot by participants of different cultural groups, and find a solution to it;
c. study how a robot greeting using language and gestures that belong to a certain culture is perceived by humans with different cultural backgrounds, in terms of acceptance and discomfort;
d. create a model for culture dependent greeting selection, so that a humanoid robot can choose the most appropriate greeting depending on the context, and even adapt to the rules of a different culture.

This research has been carried out in Japan, Egypt and Germany. Furthermore, volunteers from many other countries participated in the experiments. The humanoid robots used in the experiment are KOBIAN, its enhanced version KOBIAN-R, and ARMAR-IIIb.

The thesis is divided in two main parts, one for the face (chapters 3 and 4) and one for the body (chapters 5 and 6). Each part contains different kind of studies and experiments. chapters 1 and 2 introduce the research field and the robots; Chapter 7 concludes the work.

The thesis is laid out as follows:

Chapter 1 introduces the research background with a detailed history of cultural differences in human-robot interaction, not only between East and West but also taking Middle East into account. Then, non-verbal communication is introduced. An overview of both aspects (facial expressions and gestures) is presented, in the terms of human-human communication as well as of the state of the art in robotics. Finally, the purpose of this research will be restated.

Chapter 2 introduces the humanoid robots KOBIAN, its predecessors WE-4II and WABIAN, its evolution KOBIAN-R with the new head configuration, and finally the German robot ARMAR-IIIb.

Chapter 3 presents the emotional and communication acts model, based on Plutchik's model and Poggi's works; the mapping from human face muscles to KOBIAN-R's face, which is based on Ekman's studies; and how the generation of facial expression is achieved through classifiers and training data based on studies by Smith and Scott. Preliminary tests done to assess the generator are shown, including evaluation of random expressions.

Chapter 4 is the continuation of the work of Chapter 3 and it is structured into different subsections:

a. The expansion of the generator, capable of producing asymmetrical expressions, is described here. An overview on the state of the art regarding asymmetry in facial expressions is also included, and finally results of the experiments are shown.
b. As a recognition gap among different cultural group is detected, the generator is used in different modes for generating "Western" or "Japanese" facial expressions. The work of illustrators and cartoonists, whose drawings were used for extracting significant data, is shown. Experimental results of preferences in facial expressions and display of symbols on the robotic face are analysed, finally prompting the development of new hardware necessary for such display.
c. A further study on context-dependent evaluation is then described. It is composed of three different surveys which assess KOBIAN-R's non-verbal communication abilities through pairing its facial expressions to congruent, incongruent or ambiguous sentences. The impression that human participants feel and the robot's degree of credibility change in a way that is similar to humans.

Chapter 5 describes two cross-cultural experiments. The reactions of human subjects involved in a simulated video conference with KOBIAN were observed. The subjects were either Japanese or Egyptian and the robot was greeting and speaking either in Japanese or in Arabic. The investigation, whose data were gathered through Bartneck's questionnaires, was focused on understanding whether Egyptians and Japanese prefer a robot adapted to their own culture, and even feel symptoms of discomfort when interacting with a "foreign robot".

Chapter 6 represents the continuation of Chapter 5: as discomfort was measured in the previous experiment in case of a robot perceived as foreign, a model for culture-dependent greeting selection was made. The model features a mapping that can evolve from one culture to another. A survey of the state of the art in sociology is done in the field of greetings and politeness, including the works of Brown and Levinson. In this chapter, the implementation of the gestures on ARMAR-IIIb and the results are also described.

Chapter 7 concludes the thesis. Results are restated and broader considerations are made in this section. Finally, future works are discussed.

Appendix A and Appendix B contain respectively some examples of facial expressions produced from random input vectors, and the full questionnaires used in the experiments described in chapters 5 and 6.

In conclusion, the following results were achieved. A generator of facial expressions was designed for the humanoid robot KOBIAN-R. The system can produce different results for different cultures. Results of experiments confirmed the existence of a gap in recognition and of a bias in preferences depending on cultural group; moreover, the use of Japanese comic-style symbols displayed on the robotic face was found to be a way of making recognition easier for Japanese people. On the other hand, experiments of greeting interaction with KOBIAN suggested the existence of difference in perception between Egyptians and Japanese people, experiencing symptoms of discomfort when interacting with a robot considered foreign. The greeting selection model presented in this thesis can improve the interaction experience.

Through the careful consideration of the culture-related studies and the development of culture-specific customisation described in this thesis, it is possible to make humanoid robots be perceived as more familiar to humans, and thus improve their acceptance. As a result, not only the robots used in this research, but the whole field of humanoid robotics can take advantage of the concepts and of the models created in this thesis, starting thinking of robots as "culture-specific" machines, which need to be "localised" as any other technological device. Interaction mechanisms described in The Media Equation by Reeves should be considered, as factors like appearance and politeness are proved to play a role in human acceptance of a machine. Finally, the ability of the robots to switch between different culture modes can make tomorrow's robots be able to overcome our own cultural barriers and help us making also human-human communication easier in a global society.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Culture differences in human robot interaction

1.1.1 West-East

In the near future humanoid robots are expected to play a major role in the society, interacting and communicating with humans to help them in their work and daily life. Typical examples are serving as assistive robots for the elderly, or serving as companion robots. Assistive and personal robots are a class of devices whose design should be especially driven by considerations about human acceptance. Studies like Flandorfer's provide an interesting insight about the acceptance towards assistive robot for the elderly, underlining the importance of socio-demographic factors, including the cultural background of the human subjects [1].

Several studies have also shown that the cultural background affects the attribution of some form of personality to the robots [2], as well as the degree of anthropomorphism [3] and expectations and preferences about their role in the society and what they should look like [4], [5]. The idea that acceptance of robots depends on the culture is a very sensitive issue: Asimov was the first to introduce the Frankenstein complex [6], which describes the anxiety that people feel towards robots. The complex derives from the novel "Frankenstein; or, The Modern Prometheus" by Mary Shelley (1818) which at the beginning of the 19th Century expressed the fear that common people had for the technology, with technological creatures seen as a threat to humankind. According to the traditional view in literature, such anxiety is in part caused by popular fictional stories in which robots have negative connotations (e.g. Frankenstein, Terminator), at least in Western countries. This complex seems to be absent in Japan, where robots are viewed more like helpers or heroes. One possible explanation [7] of the latter fact lies in the Japanese animistic conception of religion, that ascribes souls to all living and non-living objects. While in Japanese mentality living beings, objects and gods are all parts of a whole picture, in the Western world, also because of Christianity, there is a strong distinction between the natural and the artificial [8]. As a matter of fact, differences between East and West in cognition, due to differing ecologies, social structures, philosophies, and educational systems, trace back to ancient Greece and China [9]. Drawing from these considerations, Kaplan suggested that "...in the Western world machines are very important for understanding what we are. We think of ourselves by analogy with the way machines work. But at the same time, technological progress challenges our specificity. That is why we can at the same time be fascinated and afraid when confronted with new machines. In Japan, in contrast, machines do not seem to affect human specificity..." [10].

However, stereotypes are not always true. For example, some of the oldest myths of artificial creation come from Greek culture, like for instance the myth of Pygmalion, who crafted a woman-shaped statue that eventually comes to life, after he falls in love with her. Most importantly, nothing in the myth condemns the creation of this creature [10]. Another milestone in the design and development of robots came with the discovery of Leonardo Da Vinci's journals, that contained plans for the construction of a humanoid robot [11]. Robotic heroes in science fiction are present in Western culture as well, like the "cute, personable and highly marketable robots" of Star Wars [12], and also some Japanese comics are in fact influenced by western science fiction[13].

Related studies support this more complex point of view. In contradiction to the popular belief that all Japanese are robot lovers, some results show that many of them are concerned about the emotional aspects of human-robot interaction [14] and significantly more concerned than Chinese and Dutch with the impact that robots might have on society [15]. On the other hand, new robotic assistance could obtain good acceptance in Italy, Switzerland and France [16]. Through these cross-cultural studies we also learned that US subjects are more positive towards interaction with a robot compared to Mexican [14]; that Chinese are more likely to anthropomorphize robots than US subjects [17]; that Chinese evaluate robots as being more likable, trustworthy and credible more than Germans [18]; and that differences in the anxiety towards robots can be found between people speaking different languages within the same country [19].

1.1.2 Middle East

The Middle East and Islamic culture in general are unexplored terrain in humanoid robotics. To the best of our knowledge, the only known studies of Human-Robot Interaction in the Middle East were performed by Makatchev et al [20] in Doha, Qatar, focusing on ethnicity cues, and by Riek et al [21], who found significant regional differences in overall attitudes towards Ibn Sina, an Arabian looking humanoid robot. That work, however, was only focused on the Middle East, without any comparison to subjects of other cultures or robots made for other cultures. Further insights were provided by Mavridis et al [22], taking into account religion too.

While trying to penetrate into Middle Eastern culture and study acceptance of robots, we should consider, especially in the case of humanoid robots, the implications of religious beliefs in those countries.

Iconoclasm (the anti-iconic doctrine of prohibition of depiction of symbols and religion icons) is well known worldwide, but less are aware of its reasons and implications. Depiction of living beings, either animal or human, has been avoided, especially in sacred spaces. In fact, an artist depicting an image of a living being would be considered as adopting the role of creator, which is reserved for only God. The roots of the belief can be found in Arabic language as well, as the word "bara'a" (to create) is related to "al-Bari'" (the Creator, a way God is referred to), hence al-Bari' is the only one who can shape things out of nothing [23]. Therefore, iconoclasm should be considered as a potential problem and definitely as an influencing factor on the attitude of people of Islamic countries towards humanoid robots.

However, there might be difficulties not related to iconoclasm. Technology acceptance, for instance, depends also on the country that is the maker of the product, as the culture of that country may bias some aspects of the product. This fact may lead to acceptance problems, especially regarding Information Technology [24] (this was the case in Syria with the use of computers in education: teachers believed that computers should first better fit Arabic identity and culture before spreading around [25]). As a consequence, localisation of products may be done. It is necessary to understand cultural norms of the country for ensuring technology acceptance [26], [27], and in the Middle East, where society rules are often blended with religious beliefs, this is particularly important.

In this research, Egypt was chosen as a location for the experiments described in Chapter 5, and Egyptian subjects were also involved in some of the experiments described in chapters 3 and 4.

1.2 Non-verbal communication: facial expressions

1.2.1 Channels of communication

Humans communicate through different channels: verbal and non-verbal. As humans, we use different types of non-verbal cues, such as kinesics, proxemics, haptics, and paralanguage [28]. Mehrabian [29] was the first who underlined the importance of non-verbal communication, stating that the non-verbal channel is even more important than words when the content of the communication involves emotions.

Non-verbal communication can have different functions. It can express a mental state, through the exhibition of affect displays [30], [31], cues about individuals' personality [30], [32], and hints about the current cognitive state [33], [34] as well as attitude and anxiety levels [35]. It can also express relations between people, through: regulations of conversation [31], [36], emphasisation of speech [37], emblems (standardized gestures), illustrators (voluntary gestures for clarification of the meaning of the speech) and performative function (clarification of the action is trying to be achieved) [35]. Long term relationships between people can be described by expressions of intimacy and emotional closeness [38], [39], and by expressions of dominance and agreeableness [31], [38].

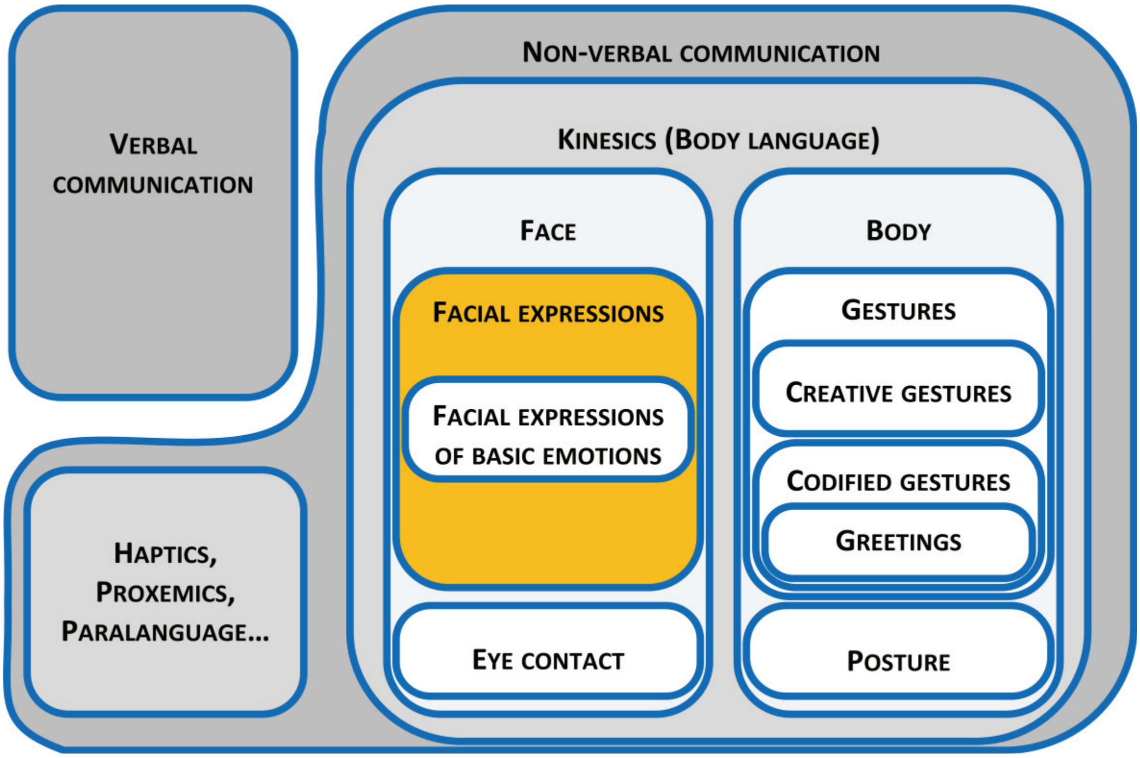


Figure 1.1 Extension of this research within communication through the face.

In this thesis, the focus is on kinesics, the study of body language. Within the scope of communication through face, in particular the study will focus on facial expressions that are not related to basic emotions, as can be seen in Figure 1.1, in orange. The reason of this choice is that it is a subfield in which culture difference are stronger [40], [41].

1.2.2 Face to face communication in virtual agents

In face to face communication, the complimentary information conveyed by facial expressions is useful for the interlocutor to understand the mental state of the speaker and even to detect lies [42].

Several attempts of artificial recreation of the use of this channel exist. One example can be found in the worldwide notorious videogame Sid Meier's Civilization. In that game it is possible to interact with virtual representations of nations leaders: this diplomacy-related communication (shown in Figure 1.2) involves bargaining and thus, changes of attitude of the virtual leader. In the most recent versions of the game, facial cues and gestures and written sentences depend on attitude and power relationship, two concepts that we will also consider. Virtual agents similar to the ones made in videogames have been developed and have brought contributions to this field.

Figure 1.2 Meeting with Stalin, leader of Russia in Sid Meier's Civilization, a strategy videogame released in 1991.

Just to name a few: virtual animated heads such as Marcos's [43], [44], or fully animated agents such as MIT autonomous conversational Kiosk [45] and Bickmore's Relational Agent [46]. Nevertheless, lack of behavioural range and expressiveness can be an obstacle for interacting with humans, and designers of such agents are aware of the challenge of rendering these agents plausible and realistic when interacting.

1.2.3 The importance of context description

One of the reasons of the lack of realism of virtual agents is characters' little depth, related to the scenario in which they live. In order to have a believable character, namely, a character that provides the illusion of life, reaction should be appropriate to the context. Typical examples of unrealism are videogames character showing no reaction to violence around them [47], or agents performing mismatching facial expression-speech pair.

This last aspect is particularly important because it may alter the measurement of expressions. As the face is considered the most important body area and channel of non-verbal communication [48], the assessment of perceived emotion must be done properly. The relative importance of context has been long debated [49], and it was proven that identical facial configurations can convey different emotions depending on the context. For instance, respectively, disgust and anger, and sadness and fear, can be easily confused depending on the scenario [50]. Cross-modal binding of voice and face was also measured through listening of neutral sentences paired with different emotional tones of voice and then matching speech with fearful and happy expressions [51], [52].

Moreover, when a face is paired to a body or voice expressing the same emotion, or when the emotional contextual scene is congruent, recognition accuracy and speed apparently increase [53]. When paired with incongruent statements, recognition decreases [54].

As verbal and non-verbal modalities are integrated by human brain, the influence of one modality is greater when the other is ambiguous or neutral [55]. Ambiguous statements are particularly interesting to evaluate, as even a smile can have different meanings depending on the social context [56]. All these findings must be considered when studying face-to-face communication.

1.2.4 Culture differences in humans

Communication between human and robots is a critical issue for the integration of humanoid robots into society. Robots should be able to understand human facial cues and display the same type of cues. However, this can be tricky even for humans, when cultural barriers are present: typically between distant cultures, such as cultures from the East and the West, facial cues have different interpretation [57]. It is known, for instance, that Japanese culture encourages the use of decoding rules (social norms that inhibit the easy understanding of emotion) [58]. Moreover, participants belonging to a certain nationality have an in-group advantage when recognising emotions expressed by members of the same cultural group [59]. Despite Ekman's [60] demonstration of existence of pan-cultural elements, differences in recognition ability do exist and were proved, among others, by Shimoda [61] and Koda [62], [63].

All these differences should be considered when implementing humanoid robots facial expressions, otherwise gaps in the accuracy in recognition between different cultural groups may happen, leading to poor acceptance and interaction for certain groups. To my best knowledge, existing studies of facial expressions in social robotics do not take care of this aspect, as robot abilities are rather limited to a small set of expressions.

1.2.5 Robots facial expressions

The focus of this research is on robots, rather than agent or humans. The involvement of the human partner in communication with robots is expected to be higher, compared to virtual agents, which have no real physical form.

Facial expressions can already be performed by a certain number of robots, including iCub [64], Albert HUBO [65], WE-4RII [66], and KOBIAN [67]. However, expression capabilities are usually limited and the small set of pre-defined patterns (typically, the six basic ones according to [40]: Happiness, Sadness, Anger, Surprise, Fear and Disgust) is one of the reason why humanoid robots are still unable to show a natural interaction. Geminoid [68], Flobi [69] and iCat [70] are other examples of recent robotic systems that are still limited to a small set, although emotion exaggeration has been investigated in the latter case. iCat and EMYS [71] are examples of robotic heads with a face that, despite retaining some human features, is inspired by an animal. In the case of EMYS, as in the case of Kismet [72], some joint movements do not have any correspondence in the human face; thus they must be interpreted separately.

Kismet was the first robot to be based on a model that could blend emotions; this kind of approach has also been attempted, in a more simplified (two-dimensional) model, on the robot FACE [73], and for full body poses on the robot Nao [74]. There is a need of extending the rigid concept of fixed patterns: we believe that this is the right direction to pursue, as a humanoid robot could be much more human-like if its expressions were not limited to the six basic emotions, but rather represent composite emotions produced by merging several parameters. This approach would be especially useful when culture should be a parameter that influences how a facial expression is made. Following the same concept, communication abilities can improve even more if expressions include moods and intentions that are not strictly defined as emotions, but that may still be communication acts which usually follow one another during a conversation. In order to take advantage of this more flexible approach and make it meaningful, a robotic head with a high number of degrees of freedom will be necessary.

1.3 Non-verbal communication: gestures

1.3.1 A world of gestures

The posture of human body can express emotion or other meanings, even when walking, but it is through the arms and especially the hands, that humans can express a wide variety of meanings. Gestures can be classified in several ways, such as the semantic content (information about the world, or about the sender), the relationship to other signals (autonomous or co-verbal) or the level of awareness of the gesture [75].

One fundamental distinction is between creative and codified gestures. The former ones can be iconic (performed when trying to depict the shape of an object) or deictic (when pointing at something) [75]. On the other hand, codified gestures have a defined meaning which varies depending on cultures. A comprehensive guide to codified gestures used in the whole world has been published in [76].

As highlighted in Figure 1.3 in orange, in this thesis research is focused greeting gestures. As codified greetings range over a very wide "world of gestures", there was the need of narrowing down the field of research, and greetings are a subfield in which cultural differences are stronger.

1.3.2 Greetings between humans

Greeting is the basic way of initiating and closing an interaction. Hoffman-Hicks [77] states that greetings function primarily as formulaic exchanges which serve to acknowledge another person's presence. We desire that robots are able to greet, same as humans. For this reason, a knowledge about the state of the art in sociology studies about greetings is necessary.

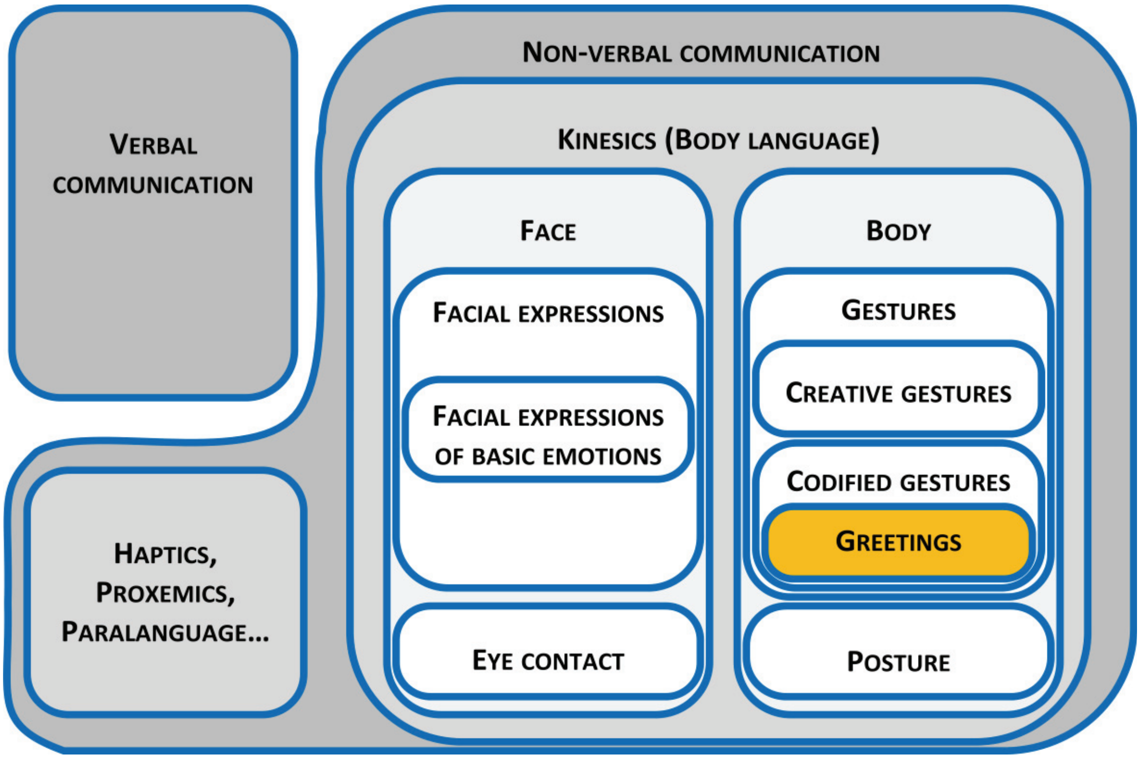


Figure 1.3 Extension of this research within communication through body gestures.

In order to define what can be classified as a greeting, the criteria of identification of greetings by Duranti [78] can be useful. A greeting has the following features:

a. Using at the beginning or closing of a social encounter. Expressions in some case can be the same.
b. Establishment of a shared perceptual field. Greetings are used as acknowledgements, if one find himself within someone's visual or auditory range.
c. Adjacency pair format. From the sequential format of greetings, mutual relationship can be understood.
d. Relative predictability of form and content. Despite there are some cases of non-predictable greeting content exchanges, as in the case of Samoan communities, usually the content is fixed.

e. Implicit establishment of a spatiotemporal unit of interaction. Greetings are a minimal proper conversation within a unit of interaction: therefore, the interaction of space and times zones of different people can be conceptualised.
f. Identification of the interlocutor as a distinct being worth recognising. In certain cultures, it is even compulsory to meet and greet after seeing each other; in some others, not being worth of the greeting is instead a marking for children and servants.

As stated previously, greetings are a form of interaction where cultural differences are evident. Depending on cultural background, there can be different rules of engagement in human-human interaction, gap in recognition of facial expressions and gestures, and chances of misunderstanding and difficulty in communication. For example, the complexity of greetings in Japanese culture may cause possible communication problems with foreigners [79]. On the other hand, Middle Eastern countries, pervaded by Islamic culture, feature some distinctive traits. For this reason, it is often suggested to study carefully customs and manners before visiting those countries [80].

There does not seem to exist a unified model of greetings in literature, but a few studies attempted a classification of greetings. Some more specific studies have been done on handshake [81]. Varieties of bowing also exist: this is why may publications guide foreigners doing business in Japan [82]

A classification of greeting was first attempted by Friedman [83] based on intimacy and commonness. The following greeting types were mentioned:

i. Smile (very common)
ii. Wave (very common)
iii. Nod (very common)
iv. Kiss on mouth (fairly common, intimate)
v. Kiss on cheek (fairly common, intimate)
vi. Hug (fairly common, intimate)
vii. Handshake (fairly common, non intimate)

viii. Pat on back (fairly common, non intimate)
ix. Rising (fairly common, non intimate)
x. Bow (uncommon)
xi. Salute (uncommon)
xii. Kiss on hand (uncommon)

Greenbaum et al [84] performed a gender related study on varieties of contact. Considered greeting types were:

i. Mutual lip kiss
ii. Face kiss
iii. Mutual face contact
iv. Handshake
v. Hand to upper body
vi. Embrace
vii. No contact

In Aisatsu no Kotoba [85], a comparative study between Germans and Japanese, the following greeting gesture types are considered:

i. Nod
ii. Bow
iii. Handshake
iv. Kiss
v. Hug
vi. Wave
vii. Raise hand
viii. No gesture

Many other contributions do not attempt to list greeting types or to classify them, but shed light on the factors that influence greeting types. In order to have a comprehensive view, the choices that influence not only gestures, but also greeting words, have been included in this study of the state of the art.

In Figure 1.4, the whole graph is shown. As Spencer-Oatey pointed out [86], authors often use the same terms with different meanings, or different terms with the same meaning. I will try to keep the terms consistent. For instance, “Context” is a word that is sometimes used, but its actual meaning is the location (private or public), that like Sugito [85] and Firth [87] mentioned, influences intimacy and greeting words; therefore I will use the word “Location” instead, and use “Context” to refer to the whole list of factors. There is also sometimes a confusion between intimacy and degree of contact. In fact, intimacy can be intended as closeness in contact, or as a close acquaintanceship: I will use just the term “Intimacy” to represent the closeness in contact, while close acquaintanceship will be described by “Social Distance”.

Intimacy is apparently influenced by Physical Distance, Eye contact [88], Gender [89], [90], Location [83] and Culture [91].

Politeness [92] [93] is a key concept in sociology: Brown and Levinson were the first to think of a formula for calculating it. Even though they did not define numerically any coefficient, they represented Politeness as a function of Power Relationship, Social Distance and a cultural factor [92]. Affect (by Slugoski [94]) can be included in this formula too, but it is usually comprehended inside Social Distance. Other factors that influence Politeness were defined by Ferguson [95]: Number of Individuals and Time since Previous Interaction. Kern [96] mentions the same factors and some others including Age, but they influence directly the choice of greeting words.

Intimacy and Politeness seem to be two concepts that stand as intermediate steps between the upstream factors and the greeting choice. Some more factors described by Li [97] influence only greeting words: Time, Regionality, Setting and Content.

Time was intended as a factor in distinguishing the use of all-time greetings, real-time greetings, seasonal greetings, introductory greetings, and ceremonial greetings. In particular, time of the day is important for the choice of words [96], [97], [87]. Setting is intended as greeting though telephone, TV or other devices, while Content refers to the use of personal greetings (either direct like "How are you?" or indirect "Your picture is beautiful, isn't it?") or non-personal ("Nice weather").

The resulting graph of the state of the art in sociology will be used in Chapter 6 to make a comprehensive greeting selection model.

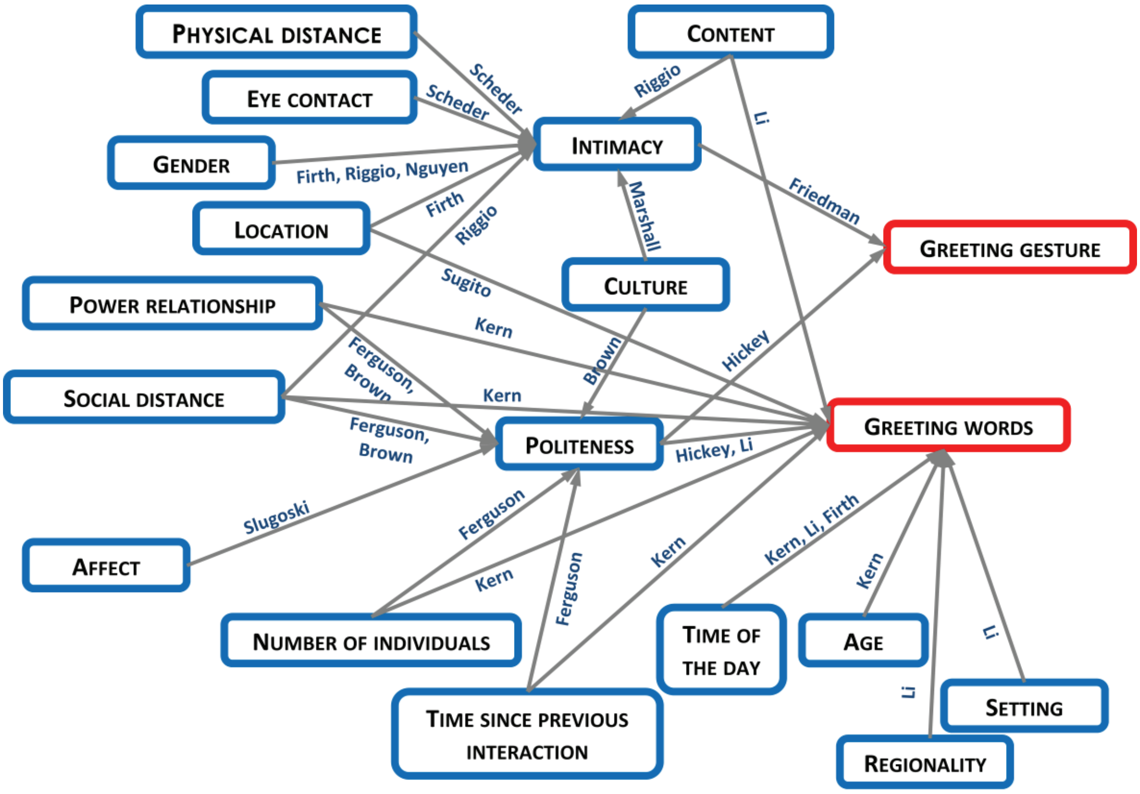


Figure 1.4 Overview of factors that influence greeting choice.
The names on the arrows indicate the authors of relevant publications

1.3.3 Robots and greetings

As robots are expected to interact and communicate with humans of different cultural background in a natural way, without generating any sense of discomfort and ensuring acceptance, it is important to study greeting interaction between robots and humans.

Some humanoid robots can perform programmed greetings. Among others, ARMAR-III [98], which met the German prime minister Angela Merkel (Figure 1.5). In that interaction, the scenario was defined in detail a priori: the position of the hand was adjusted considering Merkel's height (164 cm); the speech was defined as "Hallo Frau Merkel", and a rose was subsequently given with the other hand.

Figure 1.5 Handshake between ARMAR-IIIa and Angela Merkel.

ASIMO [99] is capable of performing a wider range of greetings: handshake, wave both hands, and bow, and can recognise such gestures among others. HRP-4 [100] and MAHRU [101] are two other examples of humanoid robots which can greet through a simple bow.

While greeting gestures have been programmed, so far only a few greeting interaction experiments with robots have been conducted to test the impression on humans. Experiments done by Yamamoto et al [102], who focused on timing, rather than on culture, and experiments featuring the social robot ApriPoco, in which Japanese, Chinese and French greetings were compared [103], [104]. However, in experiments with ApriPoco, conclusions remain unclear due to the low number of subjects and the limited number of degrees of freedom of the robot, leading to difficulties in obtaining significant data from human biological signals. Compared to those experiments, my intention is to do a more extensive study with a greater number of subjects and a human sized humanoid robot.

1.4 Outline of this research

1.4.1 Originality and purpose

As seen in the previous paragraphs, current humanoid robots, both in terms of facial expressions and body gestures, do not adapt to different cultures and just rely on pre-defined patterns. In fact, cross-cultural studies on recognition of expressions have been done before, but those studies do not take into consideration the modification of the robot, which is an original aspect of this work.

Therefore, the primary purpose of this thesis is to re-think robots design and behaviour to comply different cultures and be perceived by humans as more familiar and finally improve acceptance.

The long term idea is that customisation of robots can be pushed further. since robots don't belong to any nationality, through switching switch between different cultural and language modes, robots could facilitate human-human communication too, like C3P0 from Star Wars used to do, a "protocol robot" [105]. This abstract idea comes together with more specific customisation of robots, which is another aspect which will be discussed in the future works.

1.4.2 Practical objectives

This thesis is structured into two main flows, for the face and for the body, which end up to the same conclusions, as shown in Figure 1.6.

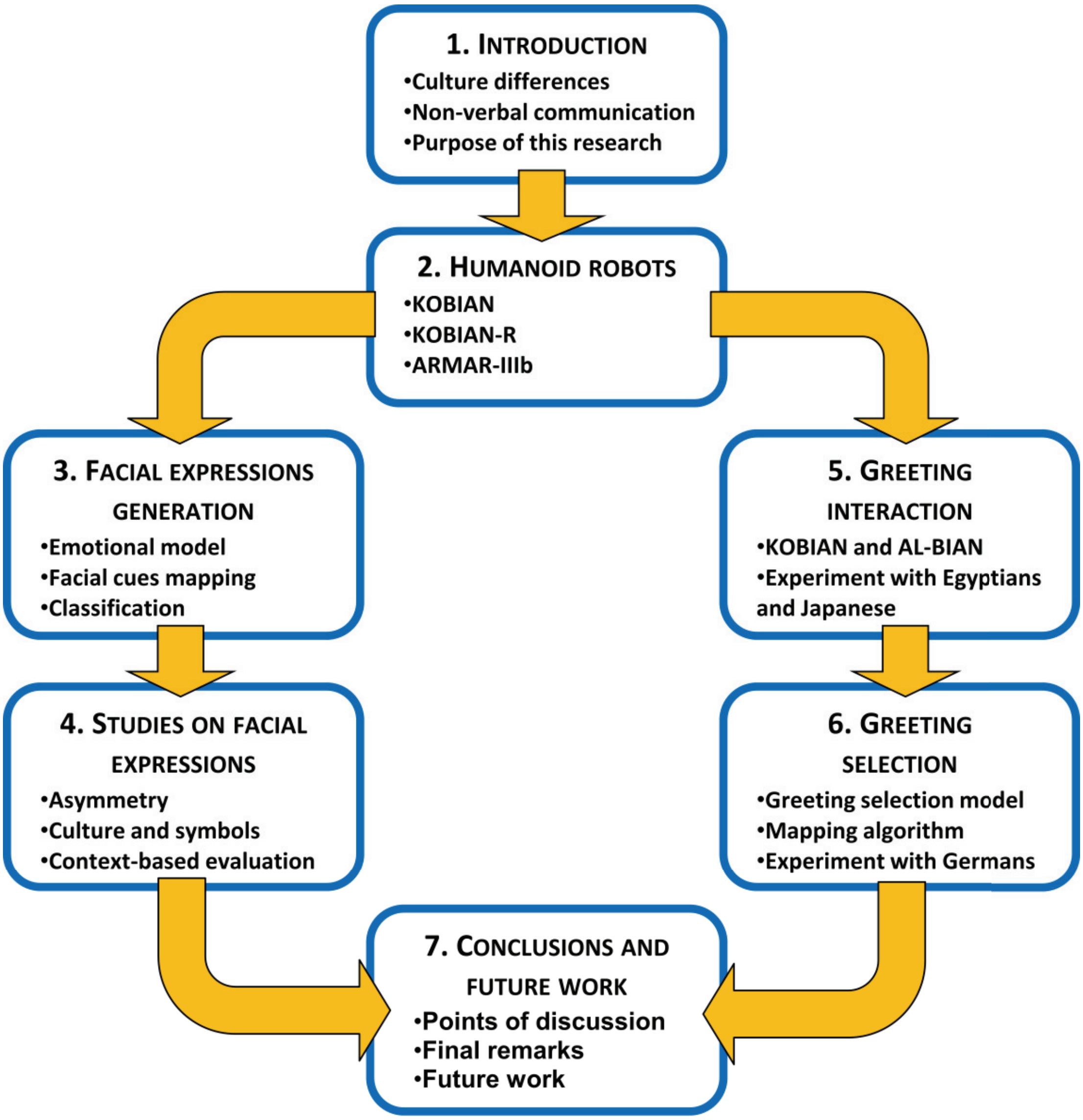


Figure 1.6 Flowchart of the thesis.

The central chapters (3, 4, 5, 6) broadly correspond to the objectives of this thesis:

a) Develop a facial expression generator and measure recognition
b) Make recognition easier for cultural groups where a recognition gap exists
c) Study how a robot greeting is perceived depending on culture
d) Create a model for culture dependent greeting selection

The reason of this structure is that for each branch (face and body), there is the need of a preliminary study (points a and c) for proving the existence of any culture-related recognition gap or discomfort. Afterwards, in case problems were highlighted by these studies, solutions must be found (points b and d).

Chapter 2

Humanoid robots

2.1 Introduction

This chapter introduces the materials, namely, the different humanoid robots that were used in these studies. KOBIAN[67] and ARMAR-IIIb [98] were used for greeting interaction, as humanoid robots that can perform gestures. KOBIAN was customised in its appearance for making a ficticial “Arabic speaking” version. ARMAR-IIIb was also customised with the addition of speakers on the back. The head of KOBIAN-R [106], the evolution of KOBIAN, was used for studies on facial expressions, for taking advance of the head’s high number of degrees of freedom.

2.2 KOBIAN

The whole body emotion expression 48-DoFs humanoid robot KOBIAN (Figure 2.1) is designed to provide support for the ADL (Activities of Daily Living) for elderly and disabled people, and to clarify the influence and effectiveness of physicality and expressivity during the interaction between human beings and robots. Humanoid robots are indeed possible candidates for being used as ADL-assistive devices, for example helping elderly people to perform activities of daily living. Besides emotion expression, KOBIAN is a robot capable of bipedal walking.

These two abilities combined together make KOBIAN potentially able, in the future, to work as assistive robot in a human environment, such as a family or a public facility.

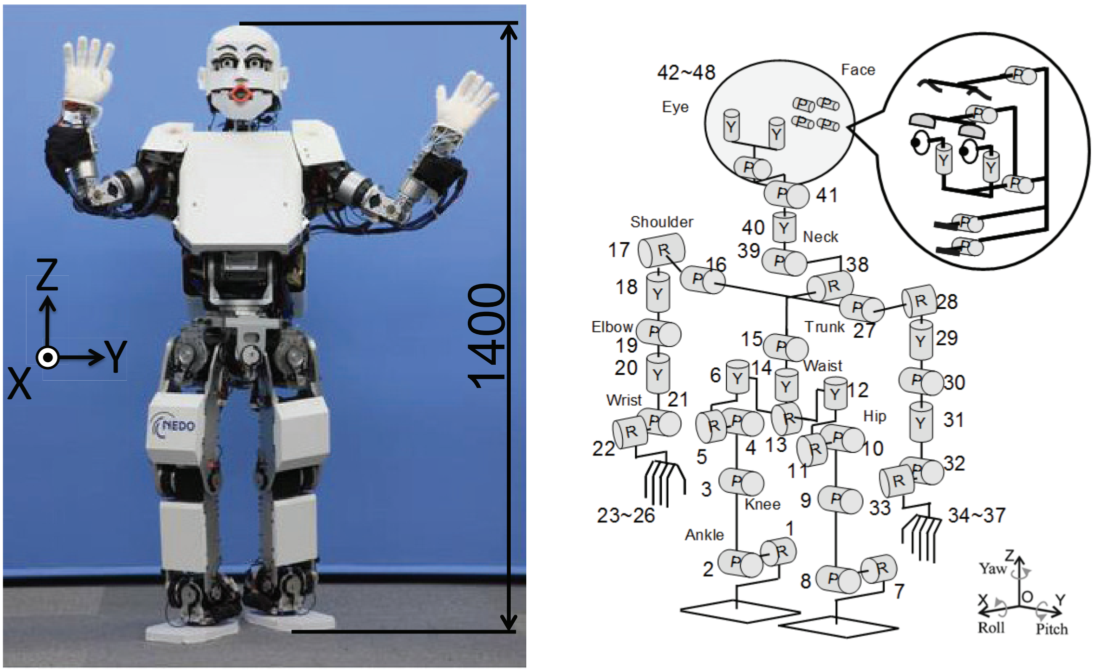


Figure 2.1 KOBIAN and its whole body degrees of freedom configuration.

Originally, KOBIAN was intended to be the combination of the humanoid robots WE-4RII [66] and WABIAN-2R [107], integrating the head of the former with body of the latter. The biped walking conditions brought some constraints regarding the size of the head and its weight, thus the number of degrees of freedom (DoF) of the head was drastically reduced to 7.

2.3 KOBIAN-R

KOBIAN-R [106]. It is the refined version of KOBIAN, with a new head. The new version of the head, thanks to the design of much downsized and lighter inner mechanisms, features a number of DoF expanded to 24, as shown in Figure 2.2 and Figure 2.3. Taking advantage of its 24 degrees of freedom, the robot has a much higher potential of expression.

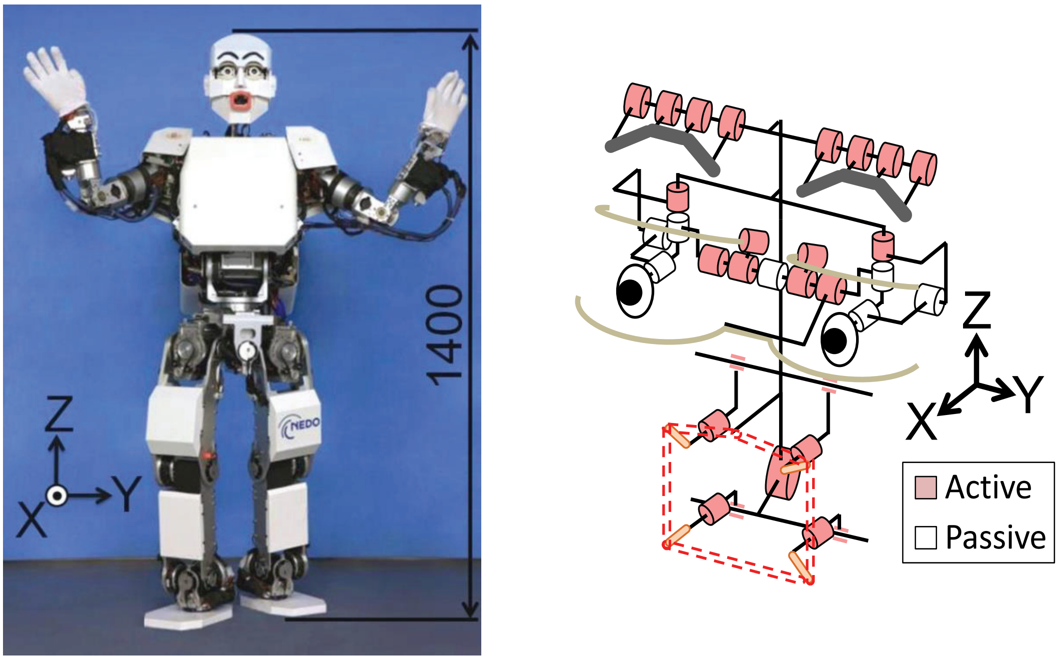


Figure 2.2 KOBIAN-R and the degrees of freedom configuration of its face.

The head has also the additional capability of changing the colour of the forehead, thanks to the use of a blue Electro Luminescence sheet behind the cover (in Japanese comics culture, blue brow represents fear). The blue colour is not visible when the Electro Luminescence is off. Despite the presence of the sheet, the forehead is thin enough to allow movement of eyebrows through magnetics.

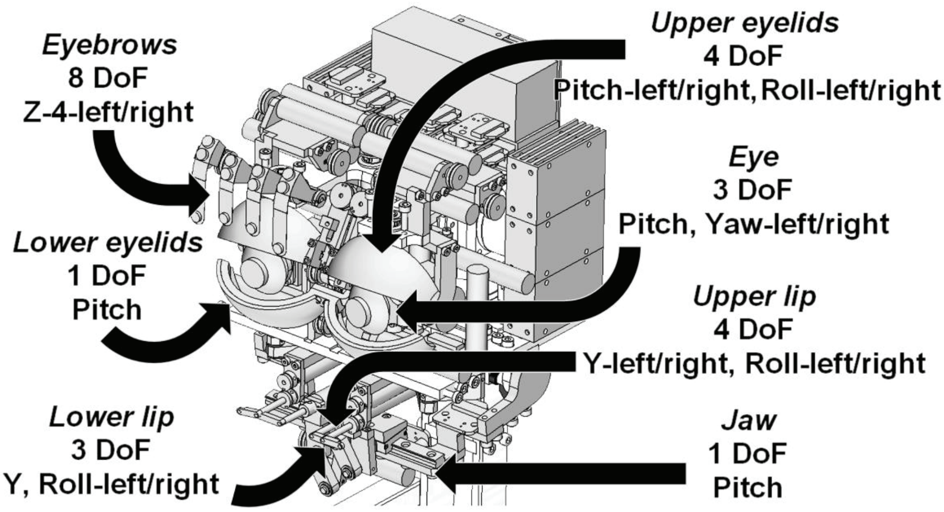


Figure 2.3 Structure of the head of KOBIAN-R.

The size of the head is 150 mm x 181 mm x 214 mm (width x depth x height), similar to a human adult female, and the weight is 1.7 kg. Lips and eyebrows are made of SEPTON [106].

2.4 ARMAR-III

ARMAR-III [98], [108] is a robot that is designed for close cooperation with humans. Unlike industrial robots, for which the primary requirements are mechanical rigidity, precision and high velocities, the most important requirements are prevention of hazards to users, a lightweight design, and a motion space that is shared with human beings. Therefore, the robot has a humanlike appearance and should have dexterity to work in human environments, and its motions should be predictable, so that also inexperienced persons are encouraged to interact with the machine.

Its predecessors were ARMAR (2000), an autonomous mobile wheel-driven platform [109], and ARMAR-II (2002): a learning and cooperative multimodal humanoid robot [110]. ARMAR-III, a fully integrated humanoid system, still wheeled, was subsequently developed in 2006 with the following features:

a. 43 DoF (7 in the head; 7 for each arm; 8 for each hand; 3 in the waist; 3 in the mobile platform) (see Figure 2.4, left picture)
b. Hands with pneumatic actuators (holding force: 2.5 kg)
c. 2 cameras in each eye
d. 6 microphones
e. Light-weight and modular upper body
f. Dimensions of the torso: 594 x 460 x 244 mm. The whole upper body is 924 mm tall (Figure 2.4, centre and right).

Two versions were built (Figure 2.5): ARMAR-IIIa, the first one, and two years later ARMAR-IIIb, a slightly modified version with different shape of the head, of the trunk and of the hands.

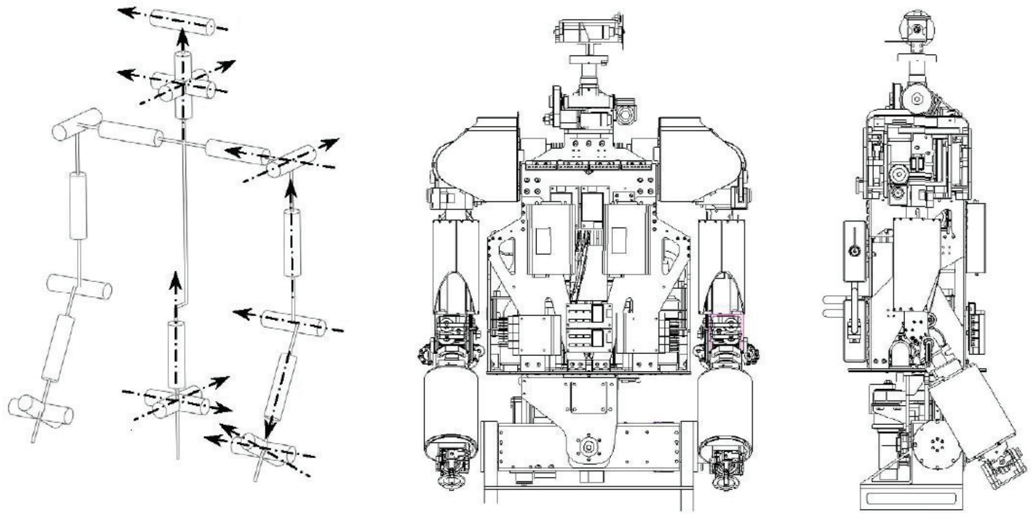


Figure 2.4 Kinematics and CAD model of the upper body of ARMAR-III.

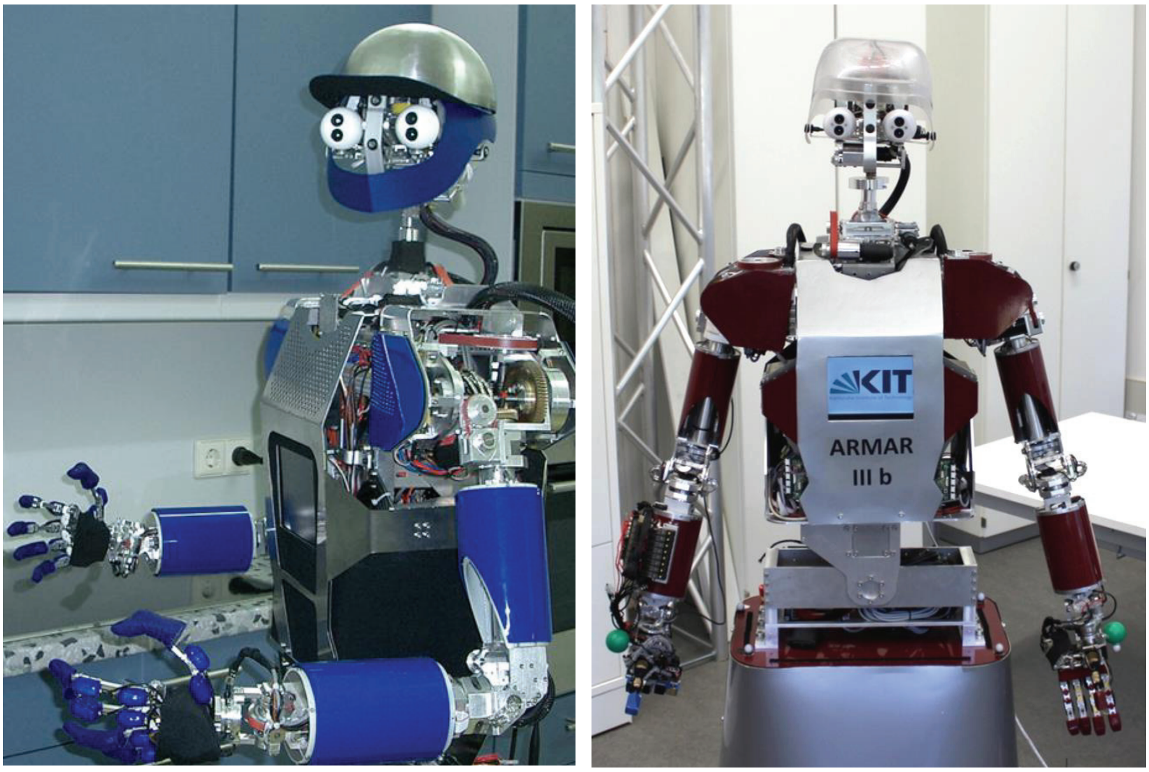


Figure 2.5 ARMAR-IIIa on the left, and ARMAR-IIIb on the right.

ARMAR-IV, a bipedal walking humanoid robot, is currently under development. ARMAR-III has sensory capabilities similar to humans because it is supposed to be able to deal with household environments, such as a kitchen, and the wide variety of objects and activities encountered in it.

The main fields of research involved in ARMAR-III are grasping and manipulation; learning from human observation; vision and exploration. The combination of these abilities make possible the execution of more complex tasks, such as the one shown in Figure 2.6.

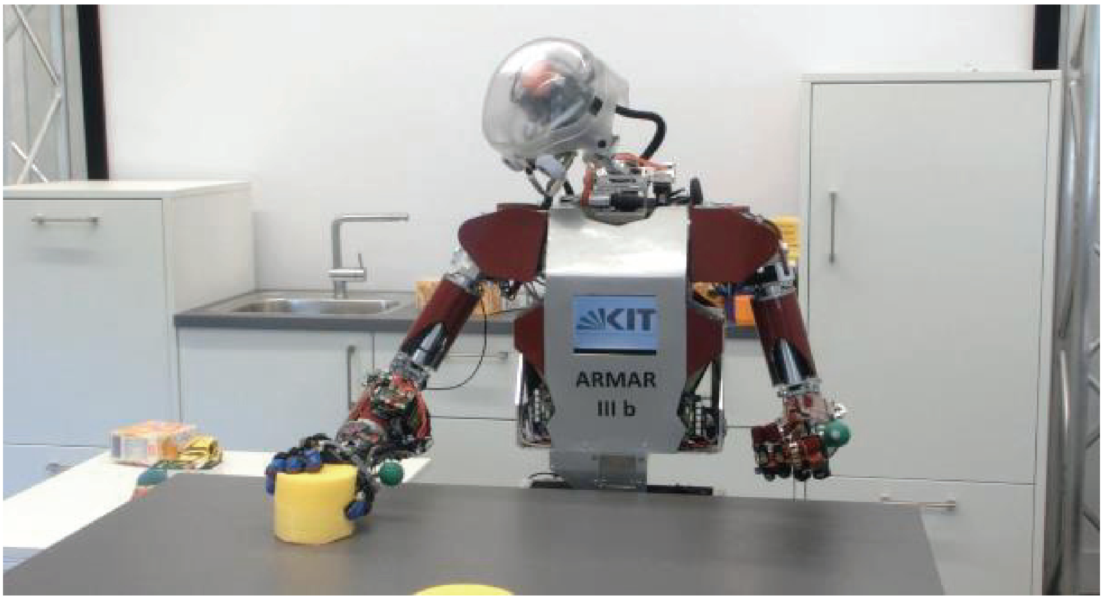


Figure 2.6 ARMAR-IIIb learning to clean the table.

Chapter 3

Facial expressions generation

3.1 Introduction

Face-to-face communication between two humans is achieved through the simultaneous use of both verbal and non-verbal communication. Humanoid robots should be able to use these two channels in order to be considered believable interlocutors.

While a few examples of robots that can already perform a certain number of facial expressions, their number is usually limited and the patterns are pre-defined. There is a need to go beyond the traditional approach of performing only the most basic expressions (fear, anger, disgust, happiness, sadness, and surprise), and rather use a parametrical approach that would make the robot able to make composite emotions. Moreover, existing humanoid robots do not take into consideration expressions that are not strictly defined as emotions, but are rather classified as communication acts (such as incomprehension or rebuke). The meaning of such complex expressions may be context-dependent: this aspect has been investigated in research on human face, but is not usually studied in deep in robotics.

In this chapter the development of a system that generates facial expressions is described. It is made for KOBIAN-R, selecting an appropriate combination of facial cues depending on inner feelings, thus resulting in a more natural and less hardwired way of displaying emotions. While the correspondence to facial cues is specific for KOBIAN-R, the emotions and communication acts model that lays behind can potentially be used on other robots. The generator is based on polynomial classifiers and on relevant studies of psychology and facial anatomy. The topics explained in this chapter have been published in [111] and [112].

Thanks to the novelty of this approach, KOBIAN-R can perform a wide range of facial expressions. Objective of this paper is proving such expressions that are adequately recognised. Context-based recognition, which is especially important in case of more complex communication acts, was also evaluated.

For making this system, we need to define an input (namely, an emotional model); an output (motor angles configuration of robotic face); and a relationship that links input and output (Figure 3.1). All of them will be explained in the following paragraphs.

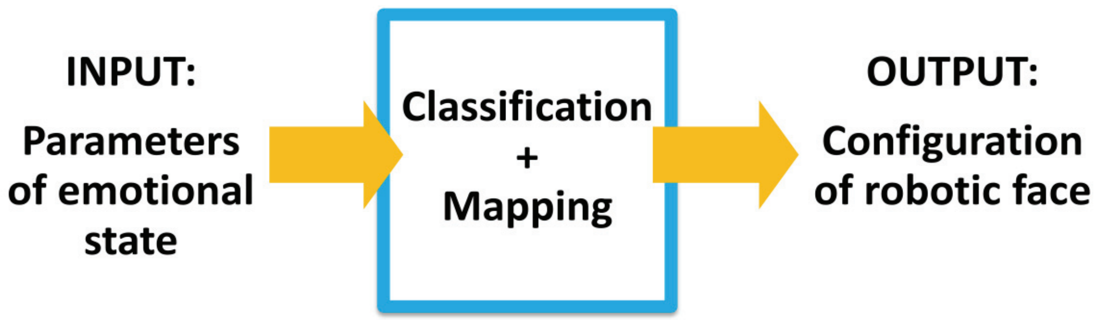


Figure 3.1 Input and output of the generation system.

3.2 Emotion and communication model

3.2.1 Communication acts

It is my intention to generate expression for more generic communication, rather than just representation of emotions. This is why we need to digress here about communication. As defined by Poggi and Pelachaud [113], the performative (namely, the actual action of interaction) of a communication act is defined by six parameters:

i. the goal (request, information or question)
ii. the interest (in whose interest)
iii. the degree of certainty
iv. the power relationship
v. the type of social encounter (formal or informal context)
vi. the affective state (the underlying basic emotion)

Some of these communication parameters should be considered for being given as input to the classifiers and therefore influence the resulting expression generation. While Goal, Interest and Type of Social Encounter would be useful in a full body dialogue system, at least Certainty, Power relationship and Affective state are useful for the present purpose, which is limited to what can be conveyed only through facial expressions and neck movement. For example, there is an obvious link between power relationship and movement of the head and eyes: when feeling superior to the person in front, usually the head tends to move upwards and eyes tend to look down [41]. The opposite is also true for feeling inferior. Degree of Certainty also influences facial cues [41].

Among the three parameters selected (Certainty, Power Relationship, Affective State), the former two do not need a more extensive study. For the Affective State, we need to use or make an emotion model, possibly using a model that is already existing in psychology and that features a wide range of emotions.

3.2.2 Emotional models

In 1980, Russell proposed a circumplex model of emotions [114]. This model defines an emotional space along with unpleasant-pleasant axis (Valence axis) and deactivation-activation axis (Arousal axis). Affective concepts fall in a circle in the following order: Pleasure (0°), Excitement (45°), Arousal (90°), Distress (135°), Displeasure (180°), Depression (225°), Sleepiness (270°), and Relaxation (315°), together with other words. The full set of 28 affect words was classified: Figure 3.2 shows direct circular scaling coordinates.

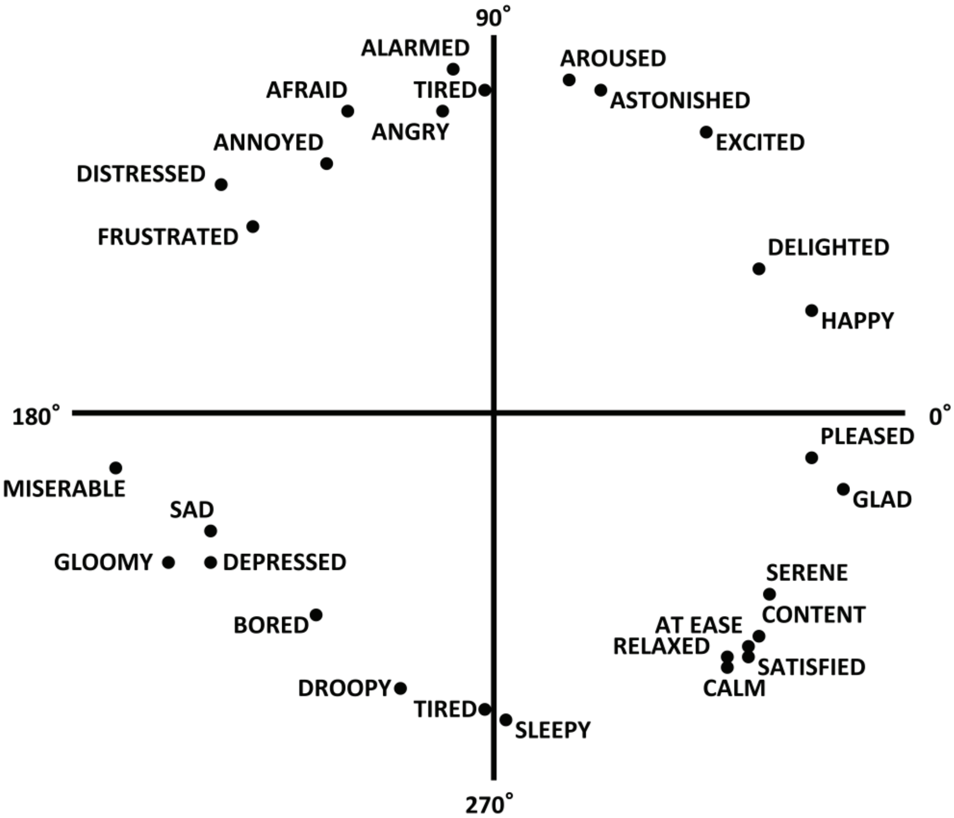


Figure 3.2 Russell's Circumplex Model.

It is possible to notice how the terms "Angry" and "Afraid" are very close in the graphical representation, and this fact makes this model unfit for being used for the current purpose.

Russell also did a cross-cultural study [115] on this model and found that results are very similar for different cultures such as Greek, Polish, Estonian, and claimed that while those results do not rule out culture-specific aspects of the human conceptualization of emotion, culture is not the sole factor of how humans conceptualize emotion.

Russell later expanded his model with Mehrabian [116] making the PAD (Pleasure, Arousal, Dominance) 3-dimensional state model. Another expansion by Fontaine [117], instead took into account Potency and Unpredictability together with Valence and Arousal. Ahn [118] shown some insights about a 3 dimensional model composed of Valence, Potency, and Arousal, and the use of asymmetry.

Scherer also proposed alternative dimensional structures for the semantic space of emotions [119]. In the framework of the componential process of definition of emotion, he defined emotion as *an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism*. In the latest (3.0) version of the Geneva Emotion Wheel [120], 20 words are organized in a wheel-like format space.

Some emotional models have been already applied to humanoid robotics. It is the case of Miwa's model (Figure 3.3) for WE-4 [121], and Breazeal's Valence-Arousal-Stance space for Kismet [72] (Figure 3.4). Both are 3-dimensional and effective in their implementation, however, they are not perfectly fitting the purpose of this research. WE-4 model is based on the 6 basic emotions and does not provide for their blending. Breazeal's model does feature the forming of composite emotions, but its complexity is not enough for the high number of DoF of KOBIAN-R's head. It is necessary a more extensive than existing 2D or 3D models.

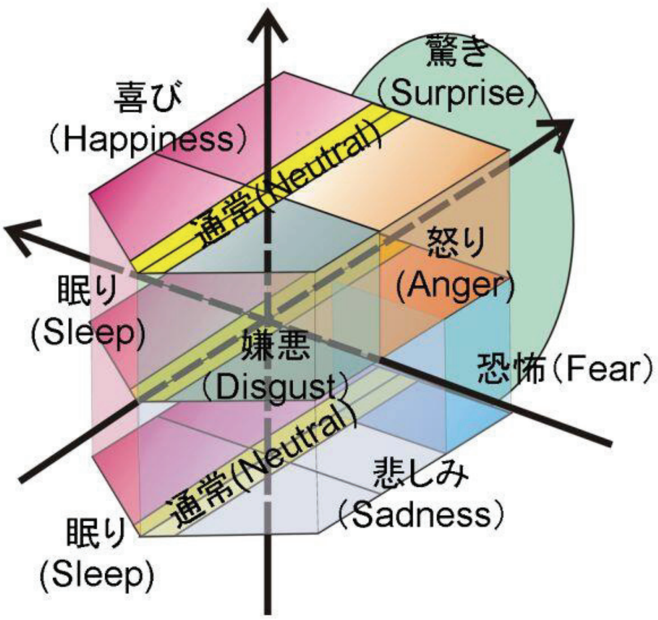


Figure 3.3 Emotion model used for the robot WE-4.

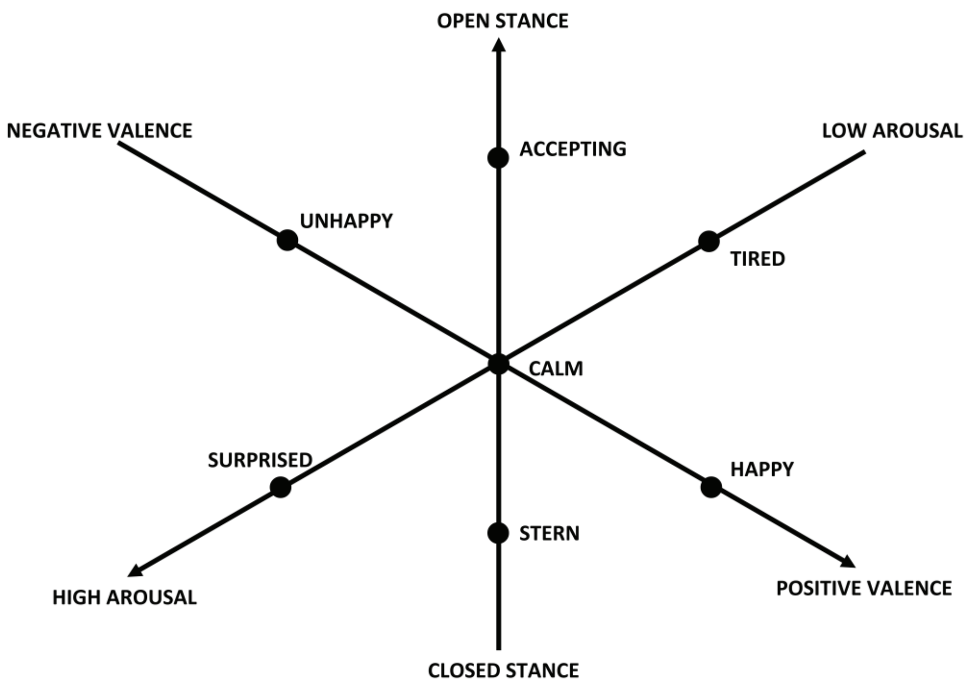


Figure 3.4 Emotion model used for the robot Kismet.

Plutchik's Wheel [122] is a model where the 8 basic emotions are opposite in pairs to each other, and where the centre corresponds to the maximum arousal, and the sides to its absence (Figure 3.5). It features a wide range of emotion labels, including secondary emotions that stand in between two primary ones. The fact that these labels are not displayed in a random order, but rather arranged in pairs, makes this model useful for the current purpose.

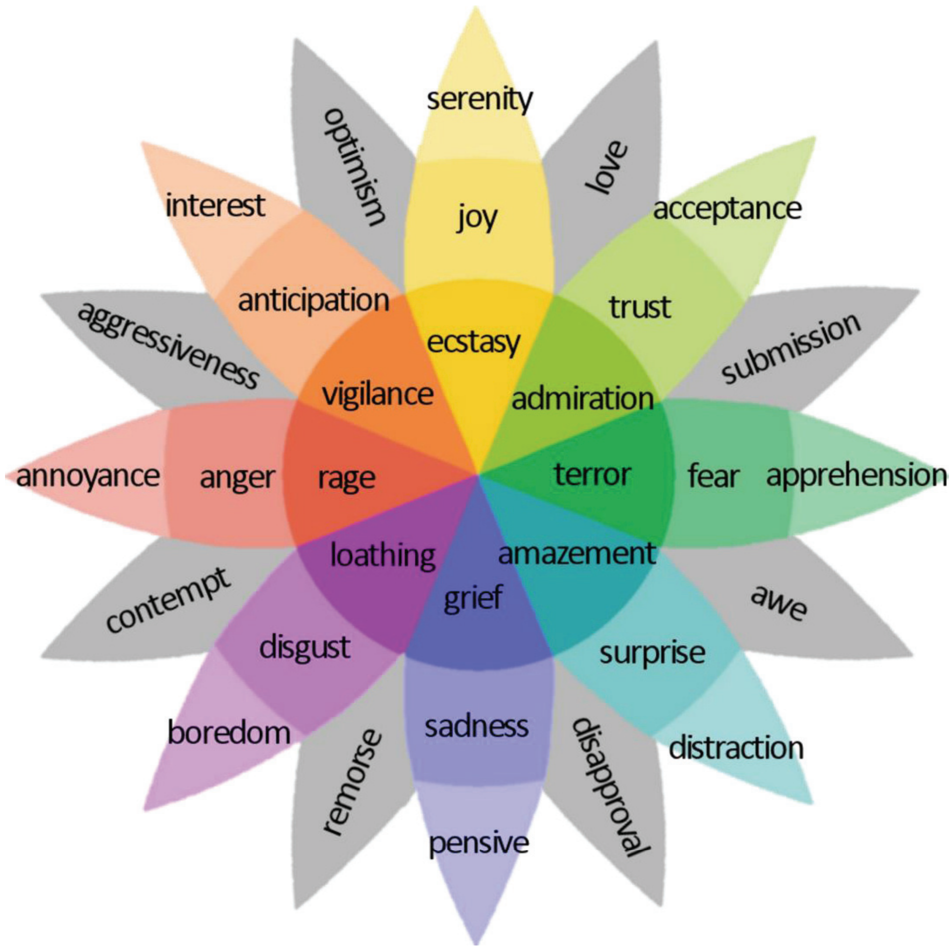


Figure 3.5 Plutchik's Wheel of Emotions, also called Plutchik's flower.

Being arranged in pairs, it is possible to convert this model to numbers using only 4 variables. I gave a name to these 4 variables (they can be seen in Figure 3.6). Arousal was not considered, as 3 levels of intensity are intrinsic within the model. Keeping a low number of variables is an important advantage in classification, and this should be kept in mind, since the chosen model will be used as input to a system that classifies such data.

3.2.3 Extended Plutchik's model

Joining the 4 variables to the 2 additional ones taken from the definition of communication act (Certainty and Power relationship, respectively measuring the certainty of comprehension of the conveyed message and the relationship with the receiver of the message), we get an extended model that consists in a 6D space that represents emotions and communication acts (Figure 3.6). The 6 basic parameters are: Mood, Stance, Temperament, Expectation, Certainty and Power relationship.

Figure 3.6 The 6 axis of the extended Plutchik's model.
Certainty and Power relationship are the the additions to the original model.

In Plutchik's model secondary emotions can be extrapolated from two adjacent branches of the wheel. In this extended model instead it is possible to span through a higher number of emotions/communication acts, which can be in between any two or more of the basic parameters. As a result, each mixed emotion or communication act is a point in the $\mathbb{R}^6$ space.

I composed an arbitrarily long list of composite emotions/communication acts, borrowing labels not only from Plutchik's secondary emotions, but also from examples mentioned by Poggi [41] and from HUMAINE Emotion Annotation and Representation Language (EARL) [123]. Numeric values have been assigned following those sources, where possible, considering that Plutchik's model has 3 degrees of intensity.

3.3 Facial cues mapping

The output of the system is the angles of the motors of the robotic face. Taking advantage of its 24 degrees of freedom, KOBIAN-R's head could able to perform a wide range of expressions if each component that can move could display a facial cue that contributes to the making of an expression.

A set of possible configurations of motor angles (resulting in a meaningful shape, appropriate for a humanoid face) was created for each facial part. We tried to maximise the number of configurations (Table 3.1, last column) while minimising the number of used variables (Table 3.1, fourth column).

In this process, the Facial Action Coding System (FACS, a procedure to systematically describe human facial expressions) Action Units (AU) [124], [125]. were used for the identifying human muscles and mapping a part of human facial cues to robotic cues.

Table 3.1 Overview of human and robotic facial cues for each facial part

Parts	DoF	N. of human AUs	N. of robotic cues variables	Set of possible configurations of the robotic facial part
Eyebrows	8	3	3	18
Eyelids	4+1	6	3	19
Eyes direction	3	6	3	23
Mouth	4+3	23	4	21
Jaw	1	5	1	4
Neck	4	9	4	-

In avatars reproducing human face [43], FACS can be used to recreate human's face muscles though direct correspondence, but this is not our case. In fact, some Action Units are indeed related to muscles that produce movements of the skin or of the nose that cannot be recreated into KOBIAN-R. For the same reason, an approach involving the extraction of real human face emotion data [126] is not possible, because it takes into consideration texture information.

In the present case, AUs are translated and adapted to the specific KOBIAN-R's face (Table 3.2), and the "robotic cues" defined in my system are fewer than AUs. This was also the case of the robots Probo [127] and EMYS [128], which faced the same problem.

Each part of the face can move in a certain way, displaying a cue. A few cues, put together, contribute to the making of an expression.

KOBIAN-R's facial parts such as eyebrows and mouth have the potential of changing into a high number of shapes. However, this potential has a downside: unnatural shapes and shapes that are dangerous for the robot (e.g. lips could break) can also be produced. For this reason, for each part of the face, a set of possible shape configurations has to be defined. In this way, only values of the motor angles that produce facial cues that are meaningful to humans, and that are safe for the robot, will be used.

Dropping unnatural shapes does not imply avoiding the use of any strictly non-human like face. One of the strengths of KOBIAN-R's face design is that it can display expressions that are exaggerations of human typical ones: such appearance is widely used in comics and actually makes recognition easier.

Reducing the number of possible outputs is not a crucial constraint anyway, since the number of possible combinations of all the facial cues is over 600'000 (multiplying values in last column of Table 3.1), without even counting the neck, whose angle values have not been discretised.

As said before, a set of possible configurations of manually chosen motor angles (resulting in a meaningful shape, appropriate for a humanoid face) was created for each facial part. I tried to maximise the number of configurations (Table 3.1, last column) while minimising the number of used variables (Table 3.1, third column). Each motor angle configuration will be described by those robotic cues variables.

In the case of eyebrows, the meaning of the 3 robotic cues used exactly correspond to AU1, 2, and 4 (Figure 3.7); but for instance, in the case of the mouth, in which AUs are present in significant number, AUs have been reduced to 8 and then paired (one the opposite of the other, such as opening/closing lips), so that just 4 variables are enough to represent robotic cues of the mouth.

The correspondence is visible in Table 3.2 and Figure 3.8. Most of the simplification involved Action Units relating to skin movements, impossible to reproduce on the robot, or multiple Action Units with very subtle differences between them, almost impossible to distinguish if implemented. A lower amount of variables is an advantage, since it will make classification of these data easier.

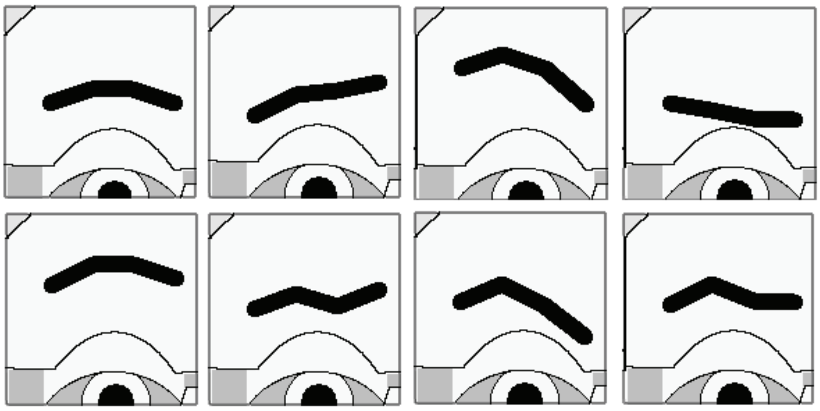


Figure 3.7 Mapping of AUs into KOBIAN-R's right eyebrow.

This set of configurations is not limited to the eight shown in the picture, but include also variations of intensity. The ones shown correspond respectively to: Neutral; AU1; AU2; AU4; AU1 + AU2; AU1 + AU4; AU2 þ+ AU4; AU1 + AU2 + AU4.

Table 3.2 Mapping of human face AUs into KOBIAN-R's cue variables

AU	Name	Robotic cues variable
	Eyebrows	
1	Inner Brow Raise	CEB(1)
2	Outer Brow Raise	CEB(2)
4	Brow Lowerer	CEB(3)
	Eyelids	
5	Upper Lid Raise	CEL(1), CEL(3)
6	Cheek Raise	CEL(2), CEL(3)
7	Lids Tight	CEL(2)
43	Eye Closure	CEL(1), CEL(3)
45	Blink	-
46	Wink	-
	Eye position	
61	Eyes Left	CEP(1)
62	Eyes Right	CEP(1)
63	Eyes Up	CEP(2)
64	Eyes Down	CEP(2)
65	Walleye	CEP(3)
66	Crosseye	CEP(3)

	Nose	
9	Nose Wrinkle	-
38	Nostrils Dilate	-
39	Nostrils Compress	-
	Mouth	
8	Lips Toward Each Other	-
10	Upper Lip Raiser	CM(3)
11	Nasolabial Furrow Deepener	-
12	Lip Corner Puller	CM(1)
13	Sharp Lip Puller	CM(1)
14	Dimpler	CM(1)
15	Lip Corner Depressor	CM(1)
16	Lower Lip Depressor	CM(4)
17	Chin Raiser	CM(4)
18	Lip Pucker	CM(2)
19	Tongue Show	-
20	Lip Stretch	CM(2)
22	Lip Funneler	CM(2)
23	Lip Tightener	CM(3), CM(4)
24	Lip Presser	-
25	Lips Part	CM(3), CM(4)
28	Lips Suck	-
32	Bite	-
33	Blow	-
34	Puff	-
35	Cheek Suck	-
36	Tongue Bulge	-
37	Lip Wipe	-
	Jaw	
26	Jaw Drop	CJ
27	Mouth Stretch	CJ
29	Jaw Thrust	-
30	Jaw Sideways	-
31	Jaw Clencher	-
	Neck	
21	Neck Tightener	-
51	Turn Left	CN(1)
52	Turn Right	CN(1)
53	Head Up	CN(2)
54	Head Down	CN(2)
55	Tilt Left	CN(3)

56	Tilt Right	CN(3)
57	Forward	CN(4)
58	Back	CN(4)

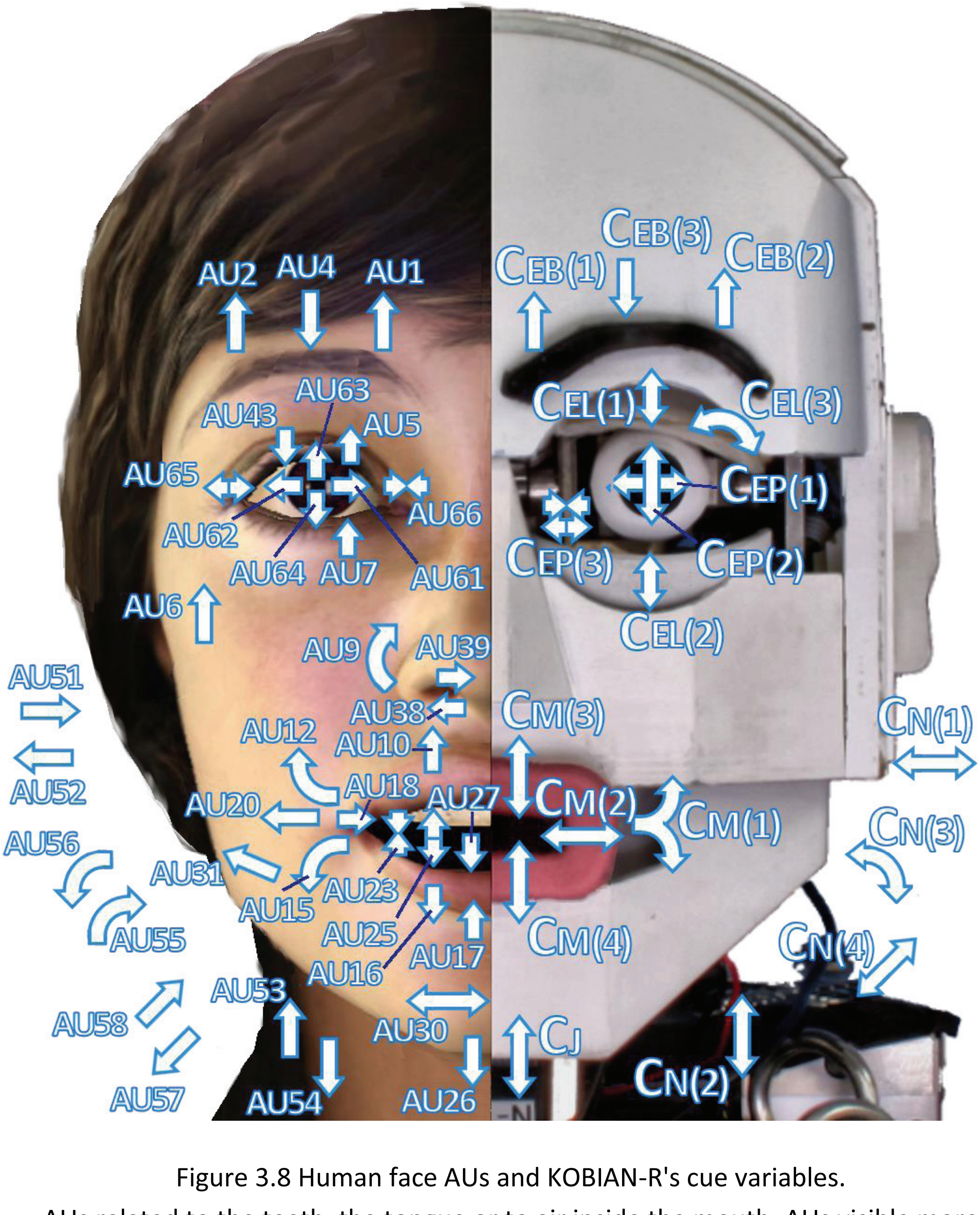


Figure 3.8 Human face AUs and KOBIAN-R's cue variables.
AUs related to the teeth, the tongue or to air inside the mouth; AUs visible more clearly from the profile, and AUs which cause changes in the appearance of the skin without significant changes in the shape of facial parts, have been left out in this picture.

3.4 Classification

At this point, while in the system the emotional parameters are given as input, the output can be chosen between a whole set of possible configurations for each facial part. Classification of input data is needed to link them with the outputs. This equals to assigning a meaning to each configuration, resulting in an "alphabet of non-lexical words".

In these regards, the studies of Poggi ([41], [129], [130]), and Ekman [131] have been used to find these correspondences. In particular, Ekman specified which AUs should appear in a face for each emotion, However, applying strictly Ekman's indications (called “categorical model", as defined by Smith and Scott [132]) is not feasible because of the difference of KOBIAN-R’s face with human face and would would not be appropriate for making a flexible system.

Conversely, I considered using a "componential approach" to the meaning of facial expressions [132], [133]. In this case, each cue is a component characterized by an exact meaning that influences the overall meaning of the expression. Smith and Scott proposed a table [132] (a portion of which can be seen in Table 3.3) that links meanings to hypothesized individual facial actions. For instance, there is a link between surprise and the action of raising eyebrows, or between pleasantness and smiling.

The training set was composed by one Neutral vector (0, 0, 0, 0, 0, 0) and four vectors for each parameters, with different degrees of intensity, as shown in Figure 3.9.

I expanded Smith and Scott’s table with information extracted from the above mentioned sources; however, the resulting table cannot cover all the combinations of secondary emotions or communication acts we are considering. As the problem gets wider, the relationships between cues and meaning can become sometimes obscure. For this reason, I decided to rely on 6 classifiers (one for each facial part) to map the correspondence with composite emotions, using the above mentioned table as training for such classifiers.

Table 3.3 A portion of Smith and Scott's table

Facial cue	**Eyebrow frown**	**Raise eyebrows**	**Raise upper eyelid**	**Raise lower eyelid**
Muscolar basis	corrugator supercilii	medial frontalis	levator papabrae superioris	orbicularis oculi
Emotion expressed / *AUs*	4	1	5	6,7
Happiness				X
Surprise		X	X	
Anger	X		X	X
Disgust	X			X
Fear	X	X	X	
Sadness	X	X		

I used degree 3 polynomial classifiers: polynomial features are added to the input dataset according to the formula of Equation 1 (where n is 6, the number of components of f, and j and k are intended as their indexes) and then classified through Fisher's linear classifier [134].

$$\sum_{j=1}^{n} f_j^{\,3} + \sum_{j=1}^{n} f_j^{\,2} + \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} f_j * f_k + \sum_{j=1}^{n} f_j \tag{1}$$

Monomials as well as combinations of 2nd order terms are constructed and then classified, which minimise the errors in the least square sense. This solution has been chosen because polynomial classifiers can map the data with very low error on the training set and produced visibly more correct outputs compared to neural networks and support vectors.

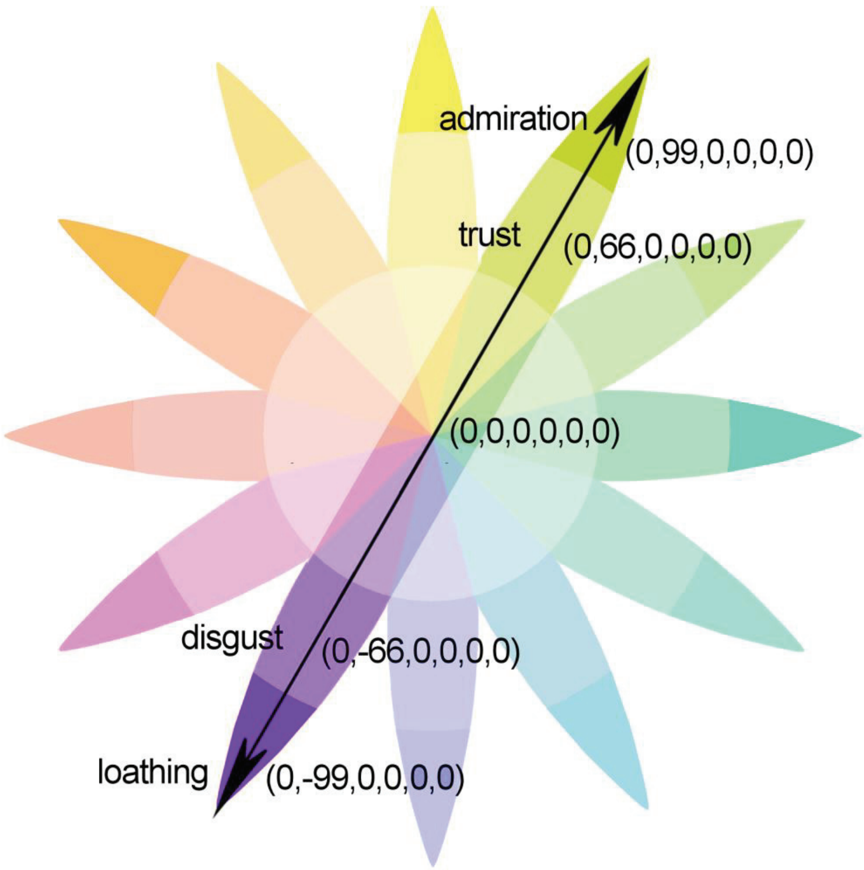


Figure 3.9 An example (the case of Stance) of vectors used in the training set. They include 4 variations of intensity of each parameter

3.5 Overview of the generator

As in Figure 3.10, the system transforms a 6 dimensional vector into a 24 one through a process of classification and mapping.. There are 6 classifiers, one for each part of the face, and one for the neck. The mathematical representation of the inner process has been defined as follows.

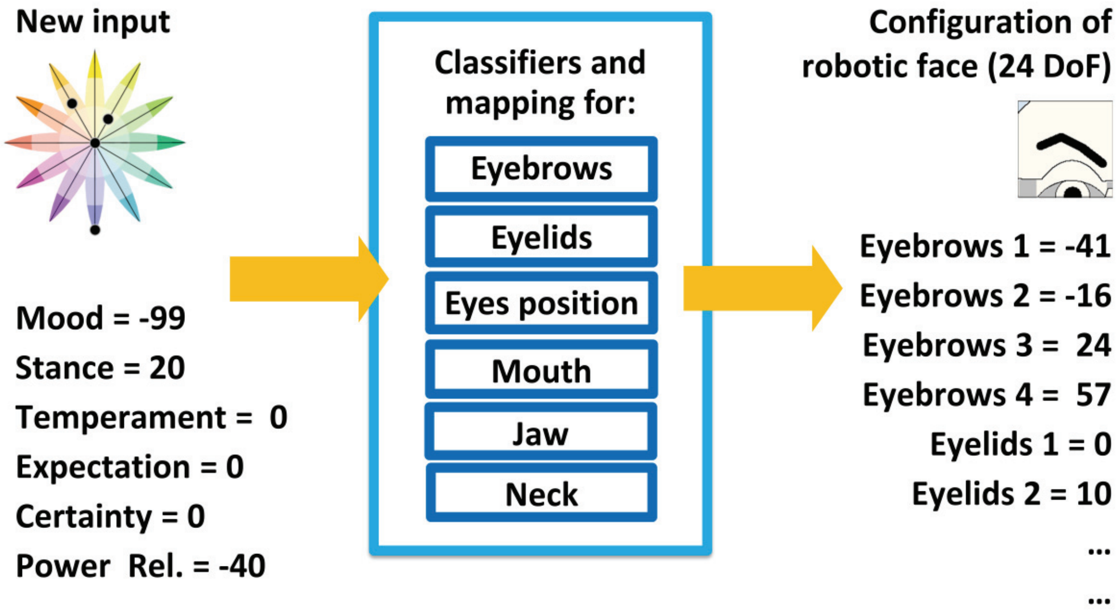


Figure 3.10 Input, output and classifiers of the generation system.

Each of the classifier is given vectors f as input. The set F containing all the f_i is a limited subset of $\mathbb{R}^6$. Each f_i is composed by the above mentioned 6 parameters $f_{i1}, \ldots, f_{i6}$, where $-99 \leq f_{ij} \leq 99$. These bounds were chosen because they are divisible by 3, so that it is easier to assign values according to 3 degrees of intensity of Plutchik's model.

The outputs of the classifiers are the robotic cues variables mentioned in Paragraph 3.3 . We call these outputs p. Values of their components range from 0 to 1 or from -1 to 1, depending on the facial part.

For each of the facial parts there is a set $C = c_1,\ldots,c_n$ of possible configurations. Each configuration is defined in the same way as the vectors p with real values ranging from 0 to 1 or from -1 to 1.

Through the use of the 1-nearest neighbour algorithm, we find, for each facial part, c^*, the best vector, which is closest to p, among the possible configurations. It represents the best configuration (defined in terms of facial cues) among a pre-defined set.

From this point, we can get the correspondent vector m^* containing real motor angles through lookup tables. Figure 3.11 shows the whole process of generation of these values for the case of eyebrows.

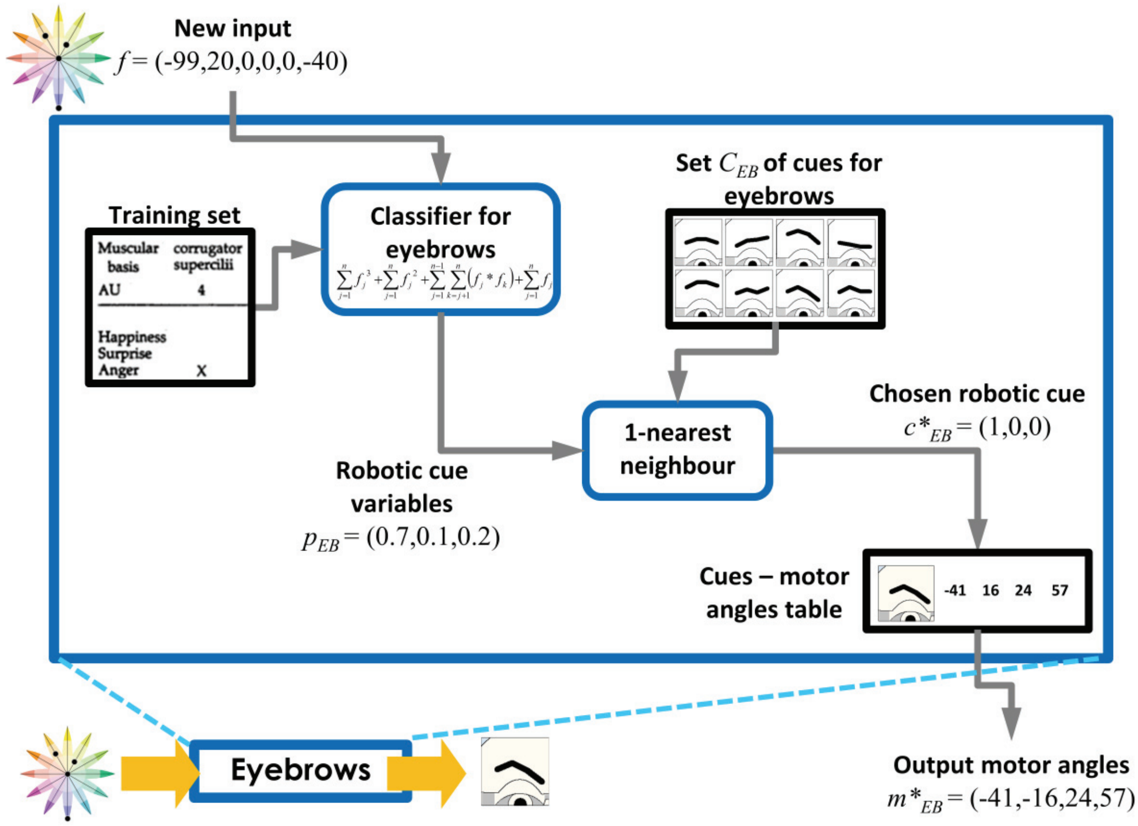


Figure 3.11 Generation of eyebrows (EB) motor angle values from a sample input with random values.

3.6 Evaluation of facial expressions

In this section, evaluation of different kind of facial expressions is described. Experiments have been done through several web surveys, inviting participants of different nationality. Two separate experiments have been made. After taking photos of the robots performing facial expressions, we made two surveys which had different goals:

- Preliminary evaluation of 12 facial expressions of complex emotions or communication acts, to confirm the effectiveness of the generation system.
- Evaluation of randomly generated expressions, a more difficult test since expressions are not labelled.

In all of these surveys, pictures of the robot (taken from the front) have been evaluated. While evaluation through direct interaction would be more interesting, in that case the number of participants would be much fewer, and a cross-cultural investigation would be much more difficult.

3.6.1 Evaluation of complex emotions and communication acts

The purpose of this experiment was to assess the generation system itself. Evaluation of expressions produced by a mix of emotions and communication acts is not expected to be easy, but the main goal was to prove that recognition rate is comparable to the one of basic emotions. Once effectiveness has been confirmed, if any culture-related difference in recognition is revealed, further studies taking culture into account can be done.

a. Participants

In this first survey, 47 volunteers (male: 28; female: 19; average age: 26.7; age standard deviation: 7.6) participated. They differed by nationality: 10 Italians, 11 Japanese, 6 South Koreans, 5 from the United States, and the rest from other countries, for a total of 25 Westerners and 22 Asians, in order to make a cross-cultural study. A group of 26 Egyptians (composed of 17 male and 9 female participants) was added later. Their average age was 28.8 (standard deviation: 5.8).

b. Experiment Setup

Subjects were asked to assign labels to 12 photos of the robotic face. Expressions were shown one at a time, next to the neutral expression as reference.

Data was divided into two sets:

- I_E, the input set containing emotions (Figure 3.12).
- I_C , the input set containing communication acts (Figure 3.13).

Each set contains exactly six expressions in order to be able to compare their recognition results with other contributions in the fields, usually based on the six basic expressions plus the neutral expression, thus giving exactly six answer choices in the survey.

The first is I_E, a collection of six expressions generated by emotions extracted directly from original Plutchik's wheel. For instance, we used, among others, Terror (0, 0, 99, 0, 0, 0), Fear (0, 0, 66, 0, 0, 0) and Surprise (0, 0, 0, 66, 0, 0) for training. In the survey, vector used for training were not evaluated, while Awe (sum of Fear and Surprise: (0, 0, 66, 66, 0, 0)) and Apprehension (low intensity Fear: (0, 0, 33, 0, 0, 0)) were evaluated.

The set I_C is a collection of six labelled facial expressions generated automatically by using input values that are not extracted from Plutchik's model. The resulting set includes communication acts that are not strictly emotions, using all 6 parameters of the extended model.

In both sets, the most basic emotions (Happiness, Sadness, Fear, Anger, Surprise and Disgust) and communication acts were not chosen because we specially wanted to measure some expressions not used for training the system. The labels of all the expressions are shown in Table 3.4. During the survey, the order of the faces was randomised.

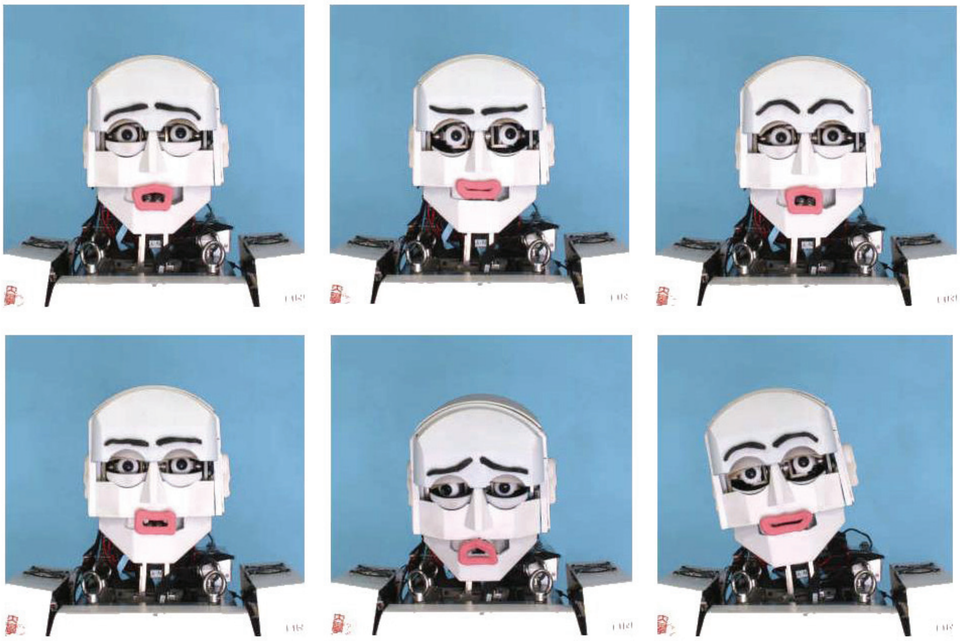


Figure 3.12 Expressions of the set I_E.

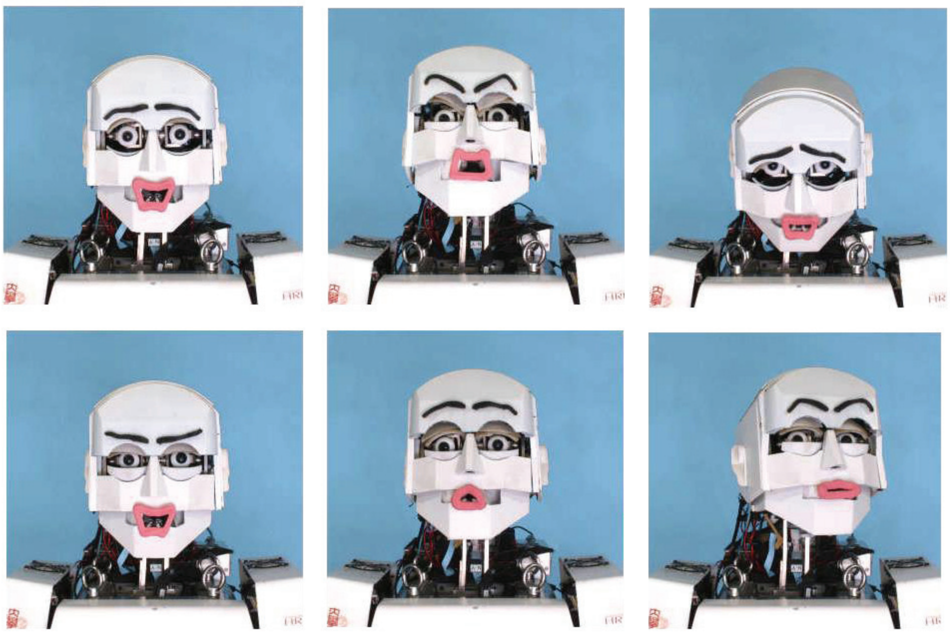


Figure 3.13 Expressions of the set I_C.

c. Results

I calculated recognition rate as the average of recall rates for each face with multiple choices. Data in Table 3.4 show that the average recognition rate of set I_E, 68.8%, is not much different from other related studies done using basic expressions, such as the results about other robots shown by Saldien et al [127], which span from 45% to 84%. Particularly significant is the comparison with the evaluation of the six basic facial expressions with KOBIAN-R's head [106], which had an average rate of 68.6%, and with KOBIAN's head [67], which had lower recognition rate. The second set has a lower average (46.4%), also due to especially low score obtained in two specific expressions (Gratitude and Pity) that confuse each other. Exchanging the two labels, rate would increase, respectively, to 56.5% and 45.7%. This would bring the average of the whole set to 55.8%. We believe that the expressions of this set, being not just emotions but rather communication acts, are generally more difficult to interpret and more sensitive to context.

Table 3.4 Results of expressions labelling

Expression	Average recognition rate	Expression	Average recognition rate
Apprehension	56.5%	Relief	50.0%
Annoyance	82.6%	Malice	34.8%
Love	73.9%	Disbelief	67.4%
Awe	56.5%	Gratitude	19.6%
Remorse	73.9%	Pity	26.1%
Hope	69.6%	Rebuke	80.4%

Table 3.5 reports the results of Egyptian participants. While the average rates are lower than the other groups, it might be due to the fact that some specific expressions hit a particularly low score. This fact begs another question, which will be discussed in Paragraph 3.7.2 : are those specific expressions difficult to understand only for Egyptians or for all cultures?

Table 3.6 summarises the results showing by average rates by country and by continent. Differences will be discussed in Paragraph 3.7 and in the next chapter.

Table 3.5 Results of expressions labelling (Egyptian participants)

Expression	Average recognition rate	Expression	Average recognition rate
Apprehension	7.7%	Relief	30.8%
Annoyance	65.4%	Malice	38.5%
Love	50.0%	Disbelief	42.3%
Awe	46.2%	Gratitude	7.7%
Remorse	73.1%	Pity	32.3%
Hope	38.5%	Rebuke	84.6%

Table 3.6 Results by country of expressions labelling

Country	N. of participants	Average recognition rate
Italy	10	70.8%
Japan	11	46.2%
Egypt	26	43.9%
Other European countries	7	61.9%
Other countries	19	56.1%
Continent	**N. of participants**	**Average recognition rate**
Europe / America / Oceania	25	67.0%
Middle East / North Africa	26	43.9%
Asia	22	47.3%

3.6.2 Evaluation of randomly generated expressions

The reason that prompted this experiment is the curiosity to investigate what can this system produce when given random values as input. In particular, I expected that human participants could still guess what kind of emotions or communication acts lie within this expression, without necessarily knowing the label of the expression.

a. Participants

The participants were the same of the previous experiment, except for the Egyptian group.

b. Experiment setup

The set I_R (shown in Figure 3.14), is a collection of 6 unlabelled facial expressions produced by random inputs. Values in F_R, the subset of F that is used to produce I_R, are randomised between -99 and 99, and then only values bigger than the threshold shown in Eq. 2 are taken, because I considered noise random values whose absolute values are smaller than the average.

$$f_{ij} = \begin{cases} f_{ij} & |f_{ij}| > AVG_j(f_{ij}) \\ 0 & |f_{ij}| \le AVG_j(f_{ij}) \end{cases} \tag{2}$$

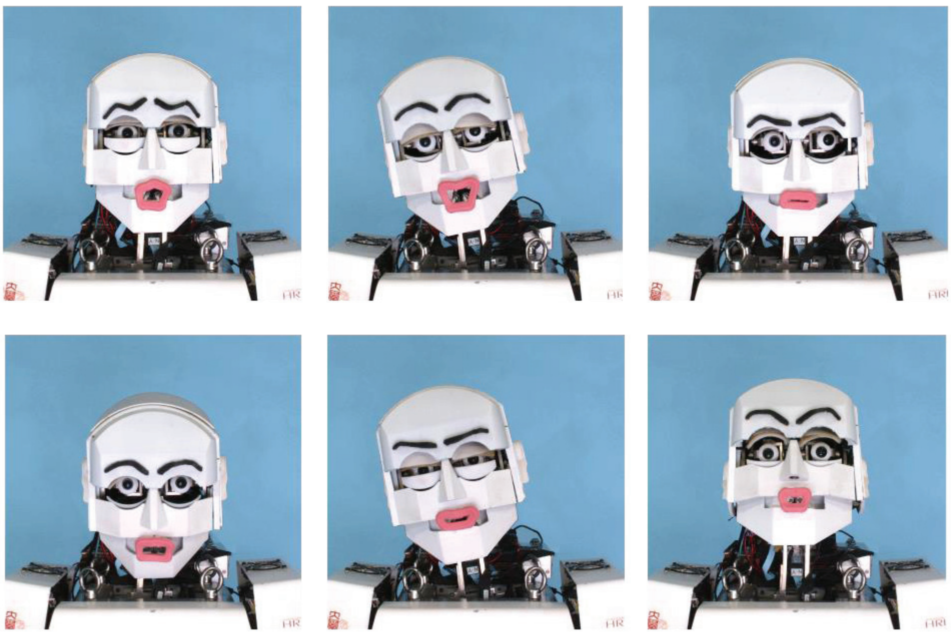


Figure 3.14 Six random expressions (set I_R).

The index i refers to the i-th expression within the set I_R; the index j refers to the parameter of the single input vector. This process results in vectors that contain 2 to 4 active parameters, which are supposed to produce ambiguous expressions, difficult to judge, but possibly more interesting to study than the other sets.

While expressions in I_E and I_C were evaluated choosing only one out of six labels, for I_R the assessment was more complicated. As it is impossible to label such expressions, there is no way of evaluating clearly if recognition is correct or not. For this reason, participants had to use semantic scales, as in Figure 3.15, for each of the parameters of the extended Plutchik's model.

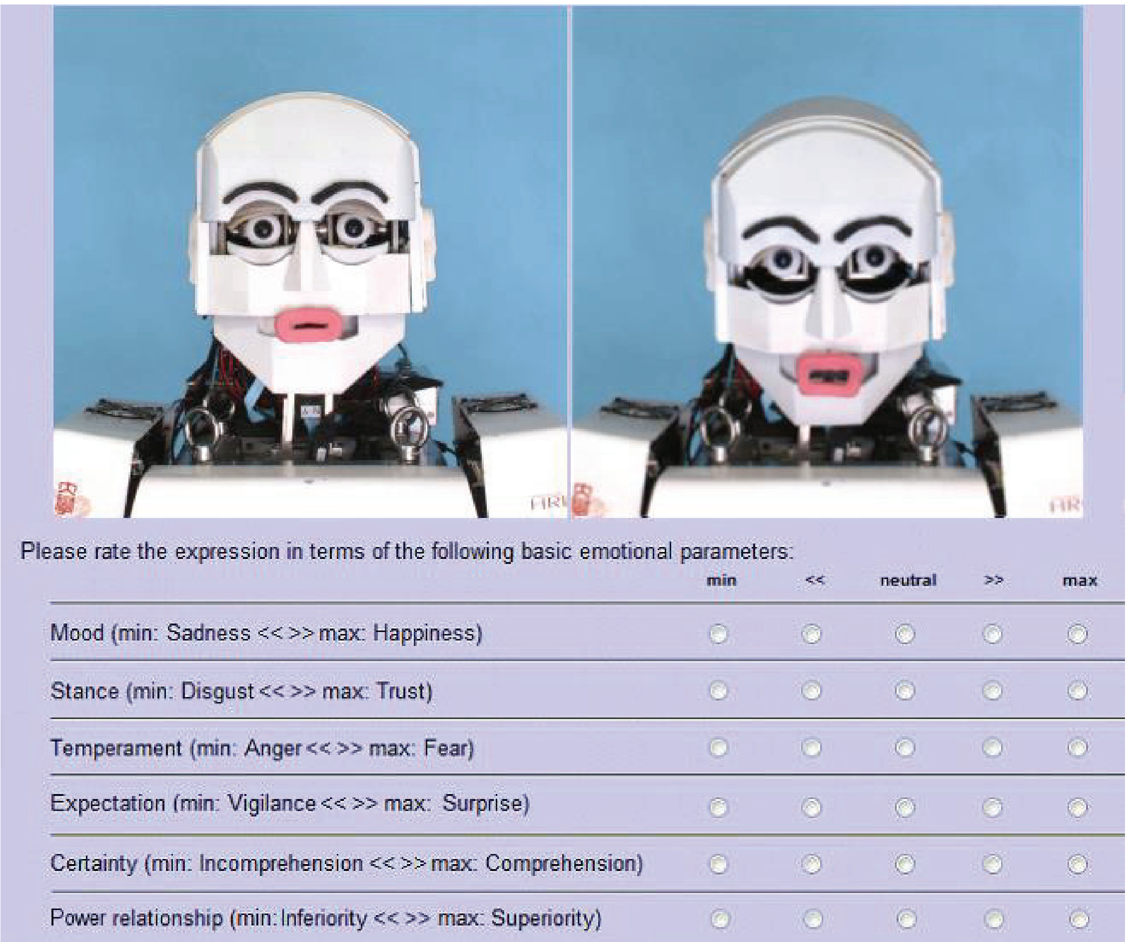


Figure 3.15 Example of questionnaire for the evaluation of unlabelled expressions.

Subsequent assessment was done in the following way:

i. We call $f_{R1},\ldots,f_{R6}$ the average results we got from each expression $f_R \in I_R$ evaluated through 5-point Semantic differential scales, one for each of the six communication act parameters.

ii. The Euclidean distance of the resulting vector of evaluated parameters $f_{R1},\ldots,f_{R6}$ from the input values $f_{i1},\ldots,f_{i6}$ that produced the facial expression gives a numerical measure of the accuracy of the recognition.

iii. The distance is then normalized dividing by the norm n, defined as in Equation 3:

$$n = \|(99,99,99,99,99,99)\| * 2 \tag{3}$$

iv. This procedure (shown in Figure 3.16) was applied also to the set I_C, so that results could be compared.

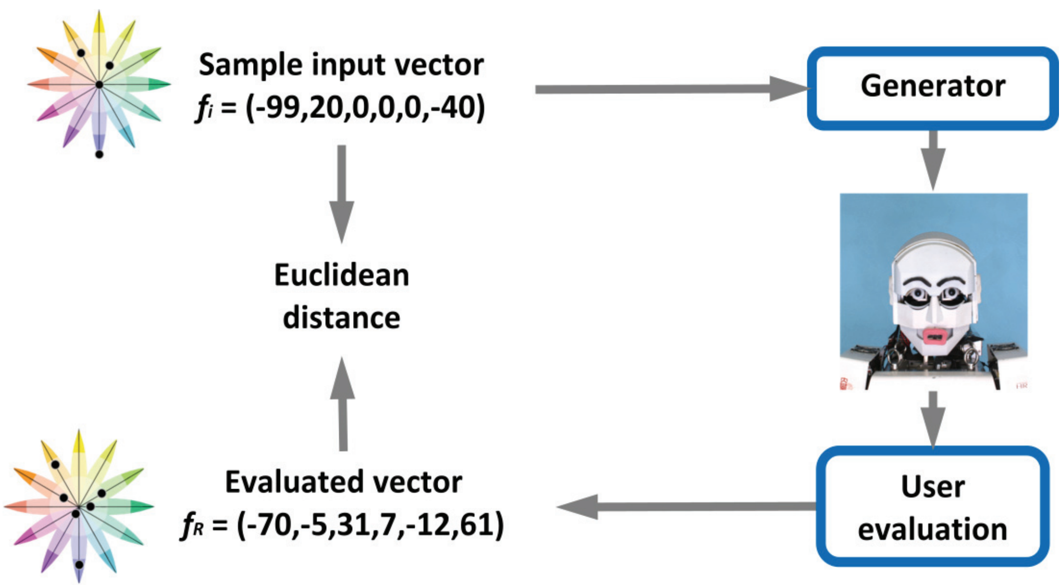


Figure 3.16 Measurement of accuracy of recognition of random expressions.

c. Results

Results related to I_R indicate that there is no significant difference between how unlabeled expressions are evaluated in terms of basic parameters, compared to labelled ones: average normalized distance is 0.222 (standard deviation: 0.045) for I_R against 0.217 (standard deviation: 0.065) for I_C. In order to understand if, in absolute terms, 0.22 is a good or a bad value, it can be argued that, answering randomly, results would span from 0.25 to 0.5. The two new parameters introduced in the extended model were evaluated separately: 0.073 for Certainty and 0.198 for Power relationship suggest that they are meaningful additions to the model. Further investigation through ANOVA on all the parameters confirms the null-hypothesis (F critical value > F) and legitimates the use of the extended model.

All the expressions of this set can be seen more in detail in Appendix A.

3.7 Discussion

Among the two sets, the average of recognition rate is lower for faces representing communication acts: the finding that the more the source emotion is complex, the more recognition will be difficult, is evident and was somehow expected. I also hypothesise that the more complexity arises, the more culture plays an important role (whereas pan-cultural elements are more likely to be revealed at basic level [60]).

The more complex interaction becomes, different ways to integrate the meaning of facial expression will be necessary. For example, communication acts could be accompanied by the use of symbolic codified gestures, as they convey an usually unique meaning [41].

3.7.1 Cultural gap: West vs East

In Table 3.6 we can see an overview of the different scores obtained dividing participants by country. Specifically, it is interesting to notice that Japan seems to have the lowest average. Expression-specific differences are also present. The most noticeable case is the expression of Awe: 100% recognition by Italians against an average of 36.4% by Asians (China, Japan, Korea and Indonesia). In a West/East comparison, analysis through t-test on the first two sets show that there is a gap in recognition ($p < .05$; *critical t* = 2.08). Average rate for Westerners is 67.0%; for Asians 47.3%. For this reason, I decided to put effort in the making of facial expressions in a way that are easy to understand for Japanese people. Such study is described in Paragraph 4.3 .

3.7.2 Recognition by Egyptians and cross-cultural patterns

Results for Egyptian participants were in general quite poor, but the high standard deviation (23.3) among the 12 expressions suggests that the problem is specific of some expressions which are not suitable for their culture. In Figure 3.17, it is possible to notice that the "pattern" of recognition rates for each cultural group has a different shape. Three particular cases are highlighted, as their standard deviation is quite high (in some cases, more than 20%) between the 3 groups.

Awe is a noticeable case of cultural gap: in general, the score of the Western group is significantly higher than the other groups. The expression "Apprehension" is one case in which Egyptians stand out from the other groups: Egyptians scored only 7.7%, while the average recognition rate for all the other participants was 56.5%). The standard deviation between the groups for Apprehensions is the highest of all the expressions (27.5%). Malice, instead, is one case in which the recognition rate of Asians is particularly low.

In conclusion, if a gap between Western countries and East Asia exists, the gap with Egyptians is even more significant. My explanation is that, as discussed in [57], it is possible that facial cues are interpreted in a different way. While a few basic cues (such as a smile) are universal [76], more complex meanings correspond to more complex interpretations. These remarkable differences confirm that it is necessary to develop culture specific expression sets.

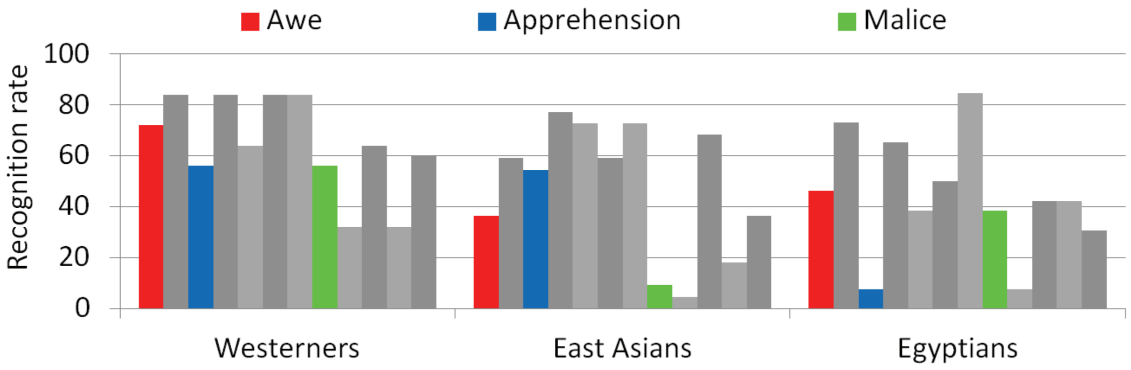


Figure 3.17 Patterns of the recognition rates of the cultural groups.
Highlighted the expressions that had a significantly different score in one of the groups, thus corresponding to a high standard deviation

3.8 Summary

In this chapter, I have proposed a system that generates facial expressions for the humanoid robot KOBIAN-R, choosing a combination of facial cues, rather than using predefined patterns for reach emotion. The parameters involved in the generation are taken from a model that describes emotions and communication acts of the robot. This model was made considering relevant studies on human anatomy and psychology.

This system was evaluated by web surveys, using photos of expressions. Results revealed that human users can effectively read in the robot's face the meaning of the expression, in a way that is comparable to the recognition of the most basic facial expressions.

Chapter 4

Studies on facial expressions

4.1 Introduction

In the previous chapter, a facial expression generation system for KOBIAN-R was developed. It is now possible to carry on additional studies. In this chapters, three topics have been investigated:

- The effect of the use of asymmetry on the conveyed emotion, and the relationship between asymmetry and culture. The contents of this paragraph have been published in [135] and [136].
- A cultural study focused on Japanese culture, given the cultural gap highlighted in the previous chapter. It was published in [136] and [137].
- A study on context by matching expressions and written sentences. It was published in [112].

While the second study is the most pertinent to the culture related theme of this thesis, the other two are still useful for investigating non-verbal communication with KOBIAN-R.

4.2 Asymmetrical expressions

4.2.1 Introduction

One aspect that could help robots being perceived as real is asymmetry. Quality of expressions can be improved taking asymmetry into account. In fact, human face is often not symmetrical over the central vertical line. Both emotional expressions and the face at rest can show signs of asymmetry. In character animation, asymmetry is an important way of making a drawn character not appear stiff and still [138]. I wanted to use asymmetry on the robot to produce expressions that look more natural, and thus are more easily recognised. In case of 3D avatars, implementation of asymmetry in a facial generator has been already attempted [118], [139].

The objective of this study was investigating the effects of asymmetry on participants' recognition comparing with symmetrical faces. Possibly also finding any correlation between facial asymmetry and culture.

The experiment consisted in two parts:

- Evaluation of the preference of participants for asymmetrical expressions compared to symmetrical ones,
- Evaluation of influence of asymmetry on a happy face.

4.2.2 Study of asymmetry

Several studies on human face have been made in the past. It has been proved that left hemiface (the left side of the face) produces more intense expressions [140]. In addition, there is a difference between genders, as males show generally more lateralisation [141]. Symmetry is also one of the factors that determine attractiveness [142]. Asymmetry has been measured, differentiating between structural asymmetry and movement asymmetry, with the latter being recognised as the primary source [143]. Relationship between valence and hemiface has also been investigated [144]. Moreover, correlation with handedness has been hypothesised [145].

My main concern, however, is not the investigation of neurobiological aspects, but rather the correlation between each of the emotions / communication acts and a particular asymmetrical facial cue. In this way it would be possible to train the system with this data, and to generate asymmetrical expressions.

In these regards, the works of Ahn et al [118], [139] are the closest to my intentions, though based on a three-dimensional avatars rather than a robot. In that case, asymmetric face was derived giving left hemiface a higher probability of negative emotion. My approach is different from Ahn's: I want to find out the facial cues that are usually activated when a certain emotion is involved, and put this data inside the generator. Therefore, I asked one professional illustrator and three amateur cartoonists to draw symmetrical and, when appropriate, asymmetrical versions of the same expression (Figure 4.1). The list of expressions to draw was taken from the 12 basic emotions and communication acts of the extended model (as in Figure 3.6), in addition to the neutral expression and the possibility for the artists to add more expression that they thought appropriate for asymmetry. As a result, the total of drawings made was more than 50 (4 samples for each expressions).

Some rules were then extracted from the drawings. Together with the artists, we hypothesised that, in an expression that is produced by a blend of two or more emotions or communication acts, the following correlations exist:

a. The presence of Anger causes lowering of one eyebrow and one-sided lower lip depressing.
b. The presence of Disgust causes sided upper lip raising and one eye aperture narrowing.
c. The presence of Incomprehension causes asymmetry in eyebrows and mouth.
d. The presence of Sadness and Inferiority causes sided lips corner tightening.
e. The presence of Superiority causes asymmetry in eyebrows and sided lips corner tightening.

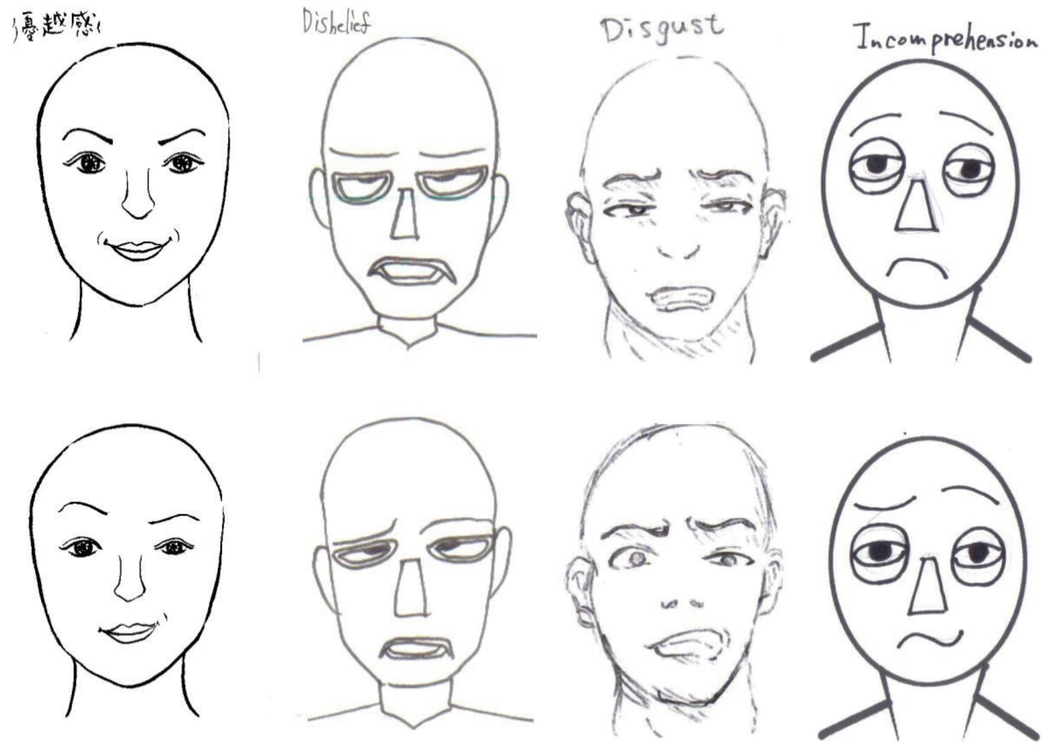


Figure 4.1 Expressions of Superiority, Disbelief, Disgust and Incomprehension drawn by artists.

It appeared that generally, asymmetry is associated with negative valence. Positive expressions, such as Happiness, seem to partially lose their positive valence, as if there were another meaning hidden beyond the happy face. We will verify these assumptions through experiment.

4.2.3 Extension of the generator

For making the generator capable of generating asymmetrical expressions, I expanded a part of the inner process. The overview is displayed in Figure 4.2.

As seen in the previous chapter, the generator takes parameters of the emotional state, gives it as input to the classifiers. The outputs of the classifiers are the robotic cues variables. These outputs are processed with the 1-nearest neighbour algorithm, in order to find c^*, the most appropriate robotic cue among the possible choices, and then corresponding m^*, containing real motor angles.

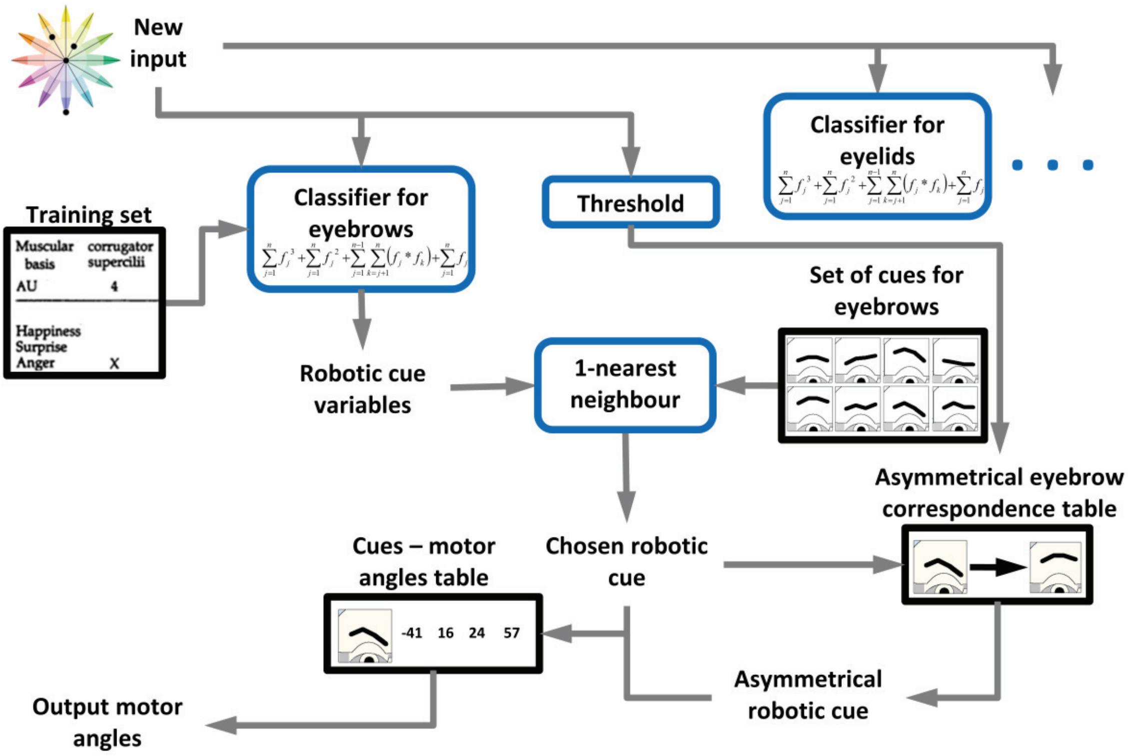


Figure 4.2 Overview of the inner process of the extended generator.

Asymmetry is managed through two thresholds θ_U (upper) and θ_L (lower). Each component f_i of the input vector gets processed: when $| f_i |$ is comprised between θ_L and θ_U, asymmetry flags for certain facial cues are activated. For each facial part that has one flag active, two different output vectors (one for each hemiface) are generated instead of one. A table associates each cue from the available set to another one, depending on the asymmetry flag. In case of more than one active flag, the one triggered by the highest f_i gets chosen (e.g. mouth can have different ways of being asymmetrical. Only one gets performed at the same time). The use of adjustable thresholds makes the system flexible. Increasing the lower threshold θ_L is a way for filtering noise: avoiding that values of f_i close to 0 have an influence on the whole face. Changing the upper threshold θ_U is meaningful for setting the system to produce asymmetrical expressions also in case of a "completely angry face" or "completely disgusted face" or for limiting asymmetry for the case of secondary emotions hidden in another expression (such as "disgust hidden in a happy face").

4.2.4 Evaluation of asymmetric expressions

a. Participants

In total, 75 volunteers (53 male; 22 female; average age: 26.9; age standard deviation: 7.4) participated in this survey. Nationality was very widespread: Western participants were 34 (7 Germans, 7 Italians, 6 South Koreans, and the rest from other countries); non-Japanese East Asian participants were 13; Japanese participants were 28.

b. Experiment setup

In the first part, the participants were asked to just choose the most appropriate of two versions (one symmetrical, the other one asymmetrical) for the following facial expressions: Disbelief, Annoyance, Disgust and Incomprehension (Figure 4.3) in a web survey as in Figure 4.4.

In the second part, they were asked to express their preference between a symmetrical and an asymmetrical version of the facial expression of Happiness. Five different asymmetrical versions were made: in each of them, one facial part at a time (among eyebrow, eyelid or mouth) had a different shape, as can be seen in Figure 4.4 in the case of the eyebrow.

Figure 4.3 Asymmetrical expressions of Disbelief, Annoyance, Disgust and Incomprehension.

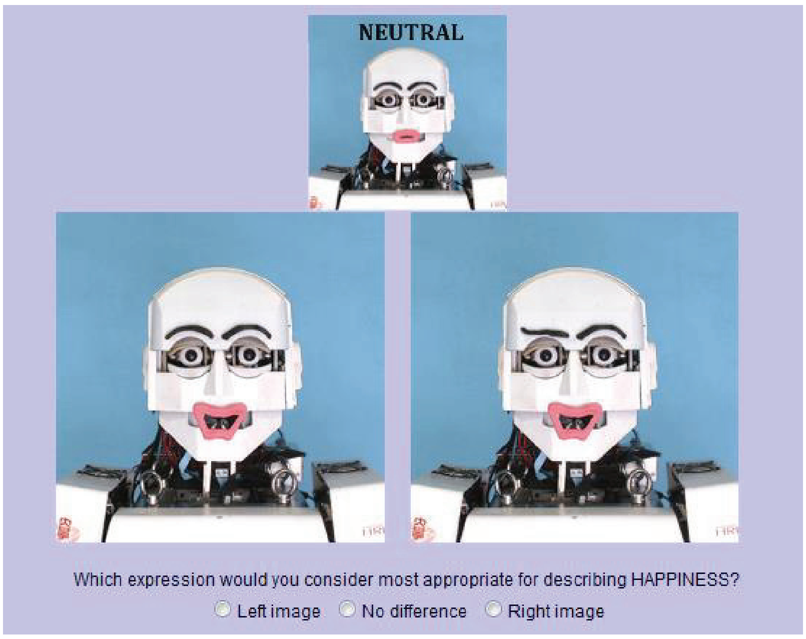


Figure 4.4 Question form for evaluating effect of asymmetry on a happy face. On the top, the neutral expressions is shown as reference.

c. Results

As shown in Table 4.1, all the asymmetrical versions of the expressions had a better recognition rate (average: 60.2% compared to 18.0% of symmetrical ones).

Table 4.1 Preferences of asymmetrical version of the same expression

	Preference for symmetrical version	No preference	Preference for asymmetrical version
Disbelief	4.00%	17.30%	78.70%
Annoyance	10.70%	37.30%	52.00%
Disgust	24.00%	20.00%	56.60%
Incomprehension	33.30%	13.30%	53.30%

The analysis of the part regarding expressions of Happiness is a little more complex. Results indeed confirm the hypothesis that asymmetry has a negative effect on perceived valence of the emotion expressed (80.2% of the preferences goes for symmetrical face, as can be seen in Table 4.2). In other words, symmetrical versions are considered more appropriate for Happiness, while asymmetrical versions hide a negative meaning. However, the attempt in finding out the association between asymmetrical facial cue and basic emotion or communication act didn't bring clear results. For the sake of clarity, I report here the emotion or communication act that had the highest number of preferences. Correlations highlighted by the survey were:

- AU4 (brow lowerer): Incomprehension
- AU43 (eye closure): Disgust
- AU10 (upper lip raiser): Surprise
- AU16 (lower lip depressor): Surprise
- AU14 (lip corner tightening): Surprise

These results are only partially matching the assumptions that were made in Paragraph 4.2.2 . In particular, all facial cues regarding mouth were misinterpreted, but the correlations of eyebrows and eyelids are correct.

Table 4.2 Evaluation of asymmetry effect on happiness

	Preference for symmetrical version	**No preference**	**Preference for asymmetrical version**
AU4 (brow lowerer)	85.30%	10.70%	4.00%
AU43 (eye closure)	90.70%	8.00%	1.30%
AU10 (upper lip raiser)	85.30%	13.30%	1.30%
AU16 (lower lip depressor)	74.70%	17.30%	8.00%
AU14 (lip corner tightening)	64.90%	27.00%	8.10%

4.2.5 Relationship between asymmetry and culture

Results obtained so far are displayed here dividing participants by nationality. From Table 4.3 and Table 4.4, it appears that there is a relation between asymmetry and negative valence of the expression, but there are no differences of perception across the groups. There are a few differences (for example, in the first study, standard deviation of the three groups is, respectively for each expression: 8.3; 24.8; 16.4; 7.6), but such data cannot prove the existence of any correlation between perception of asymmetry and culture.

Table 4.3 Preferences by nationality of different versions of the same expression

Country	Preference for symmetrical version	No preference	Preference for asymmetrical version
Western countries	22.1%	20.6%	57.4%
Asian countries w/o Japan	11.5%	15.4%	73.1%
Japan	16.1%	26.8%	57.1%

Table 4.4 Evaluation of facial asymmetry effect on happiness

Country	Preference for symmetrical version	No preference	Preference for asymmetrical version
Western countries	82.9%	13.5%	3.5%
Asian countries w/o Japan	78.2%	18.9%	2.9%
Japan	76.9%	12.3%	10.8%

4.3 Culture and symbols

4.3.1 Introduction

The objective of this part is to prove the existence and find a solution to the recognition gap discussed at the end of the previous chapter, though the following key points:

a. Create a parametric system capable of generating facial expressions in different versions, depending on the culture dependent rules the system has been trained on.
b. Generate two versions (in this study, Western and Japanese) of the same expressions and display them on the face of the humanoid robot KOBIAN-R; then ask participants belonging to several cultural groups to evaluate them.
c. Find additional ways of making recognition easier for different cultural groups.

Regarding the last point, as explained later, Japanese comic symbols were displayed, in order to find which kinds of symbols are useful for each participants group and therefore start planning their implementation.

4.3.2 Japanese artists' work

In order to make a "Japanese training set", I asked one professional illustrator and three amateur cartoonists from Japan to draw expressions for all the emotion or communication acts that are the base of Plutchik's extended model.

The table used so far for training the classifier, as it is based on sociology papers from American and European studies, will be called from now on the "Western training set".

Artists were specifically asked to consider the capabilities of KOBIAN-R's head, specifically ignoring facial cues related to nose (which KOBIAN-R can't move), skin or hair. A total of more than 50 drawings were made. After collecting the drawings, a sample of which is shown in Figure 4.5, rules for creating a Japanese training set were extracted comparing the pictures and choosing facial cues that highlight the differences with the Western set.

During this process, it has been observed that many shapes of eyebrows drawn by Japanese artists cannot be reproducible by Action Units, as they are exaggeration of real facial cues. For this reason, a different robotic cues set for eyebrows has been made (in other words, the output part of the system was also modified).

All the artists then stated that it's difficult for them to express easily recognisable emotions without the use of symbols that are common in Japanese comics and animation [146].

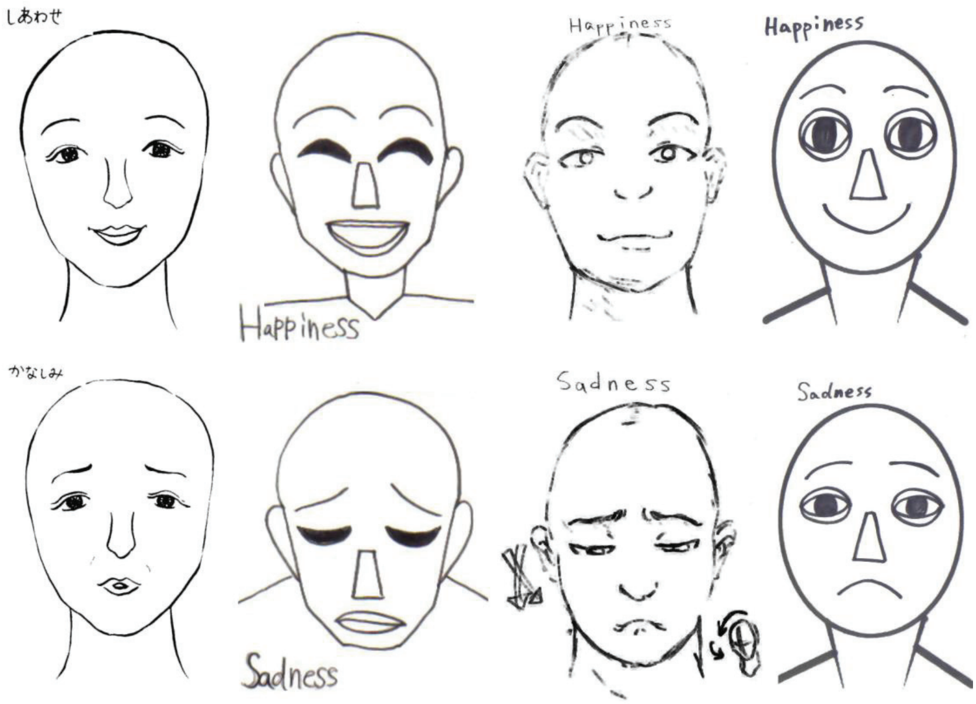


Figure 4.5 Facial expressions of Happiness and Sadness drawn by one illustrator and three cartoonists.

Some of the symbols are either usually written near the character's face, or consist in the complete change of the shape of some facial features (i.e. replacing eyes with two hearts or two stars). In such cases, their implementation is not feasible; however, a subset of these symbols, shown in Figure 4.6, could be shown on the face of the robot. This set might be expanded in the future. In the present study, the 7 symbols and the blue brow (representing fear in Japanese comics) were considered. Despite it will be possible to display symbols on the face using Electro Luminescent materials or other methods, this solution is not yet implemented. For this reason, a preliminary investigation on symbols was be done by displaying expressions containing symbols added by photo-editing.

4.3.3 Extension of the generator

Through the addition of a different training set, the generator can produce different outputs. I made a different table, similar to the previous one ("Western training set") replacing its data with the rules extracted from the artwork. In Figure 4.7 it is possible to see the overview of the generator inclusive of all face parts. The two additions are the Japanese training set, and an additional set of exaggerated cues for the eyebrows.

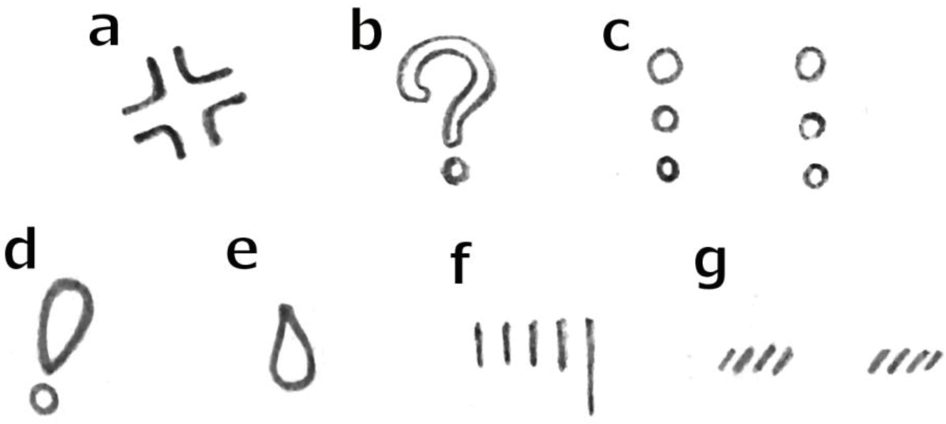


Figure 4.6 A set of symbols commonly used in Japanese comics.
a: cross-popping veins (anger); b: question mark (incomprehension); c: tears (grief); d: exclamation mark (vigilance/notice); e: sweat drop (anxiety); f: parallel vertical lines (trouble); g: red cheeks (shame).

4.3.4 Evaluation of culture dependent expressions and symbols

This experiment consisted in the evaluation of two different versions (Western/Japanese) of the same expression, all produced by the generator, in order to understand whether a customised expression, in term of different facial cues, for a certain culture, may make recognition easier.

Evaluation of expressions that display Japanese comic symbols was subsequently done, in order to understand whether these symbols are useful, and in case, which ones should be implemented on the robot.

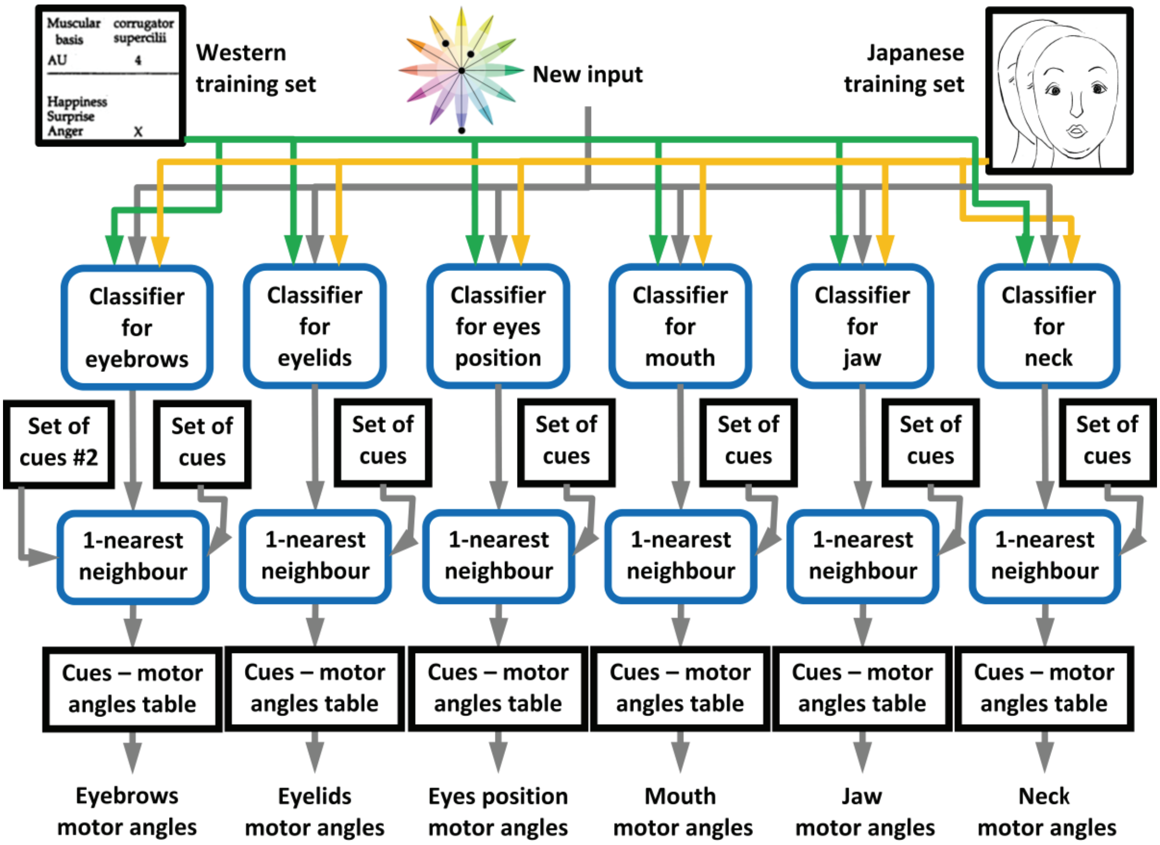


Figure 4.7 Complete overview of the generator with different training sets.

a. Participants

The participants were the same of the survey described in Paragraph 4.2.4 . In addition, the group of 26 Egyptians that took part in the test of Paragraph 3.6.1 also performed this evaluation. Participants are therefore divided in four groups: Table 4.5 and Table 4.6 contain results for all the groups.

b. Experiment setup

This survey consisted in two parts. In the first one, the participants were asked to choose between two versions of the same expression (Figure 4.8), produced by different training sets, while being able to see the neutral expression shown at the top as reference.

The expressions that have been evaluated are: Happiness, Disgust, Fear, Anger, Incomprehension, Remorse, Awe, and Malice. They were chosen either because their recognition rate by Japanese participants was particularly poor in the survey described in Paragraph 3.6.1 , or because their representation by Japanese artists was consistently different from the Western version.

In the second part, the participants had to express their preference between two versions of the same expression: one displaying a comic symbol on the forehead or on the cheeks, and one not displaying any addition to the expression produced by the generator.

In the present state of the robotic head, the whole brow can turn blue through Electro Luminescence, but it cannot display symbols. However, as it will be technically possible to implement more complex shapes, the symbols listed in Figure 4.6 were added on Japanese expressions pictures by photo-editing (Figure 4.9). If the idea is successful, it will be physically implemented.

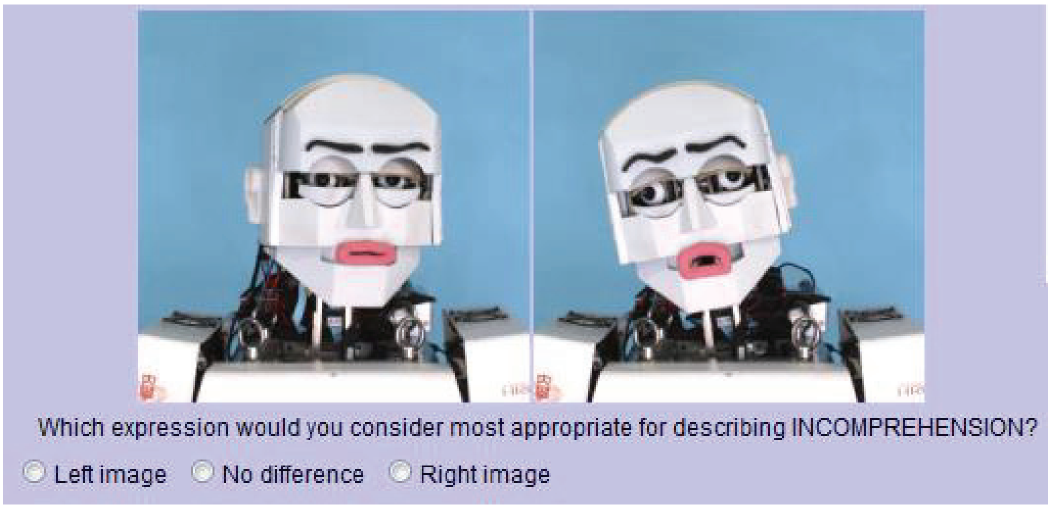


Figure 4.8 Sample of the evaluation form for culture dependent expressions.
In this case, the right one, supposed to be the "Japanese version", was chosen by 85% of Japanese participants.

Figure 4.9 Expressions containing symbols.
Top line: Fear, Anger, Incomprehension, and Grief. Bottom line: Vigilance, Anxiety, Trouble, and Shame.

c. Results

Qualitative analysis of the results in Table 4.6 shows that in-group advantage exists, as Westerners consider more appropriate expressions that were made using a Western training set, and Asians prefer expressions generated using the Japanese training set.

Table 4.5 Culture dependent expression evaluation results

Nationality	Western expression preference	No preference	Japanese expression preference
Western countries	57.4%	9.9%	32.7%
Asian countries w/o Japan	42.3%	8.7%	49.0%
Japan	41.4%	15.5%	43.1%
Egypt	55.9%	13.5%	30.5%

Although such results are significant, they could be improved through the use of Japanese comics symbols. Results in Table 4.6 and Figure 4.10 are comprehensive of Egyptian participants and show clearly the bias between the groups. Figure 4.10 shows the particular preferences of each symbol for each group. While scores are high (more than around 70%) for all symbols in case of Japanese, for the other groups (especially Westerners) preferences vary depending on the specific symbol. Usefulness of the use of symbols seems therefore to be supported by these results, especially for Japanese and other Asian participants. Further discussion follows in paragraphs 4.5.2 and 4.5.3 and in the future works.

Table 4.6 Japanese comics symbols evaluation results

Nationality	Preference of no use of Japanese symbols	No preference	Preference of use of Japanese symbols
Western countries	36.0%	32.0%	32.0%
Asian countries w/o Japan	14.4%	22.1%	63.5%
Japan	14.3%	7.9%	77.9%
Egypt	30.8%	19.8%	49.4%

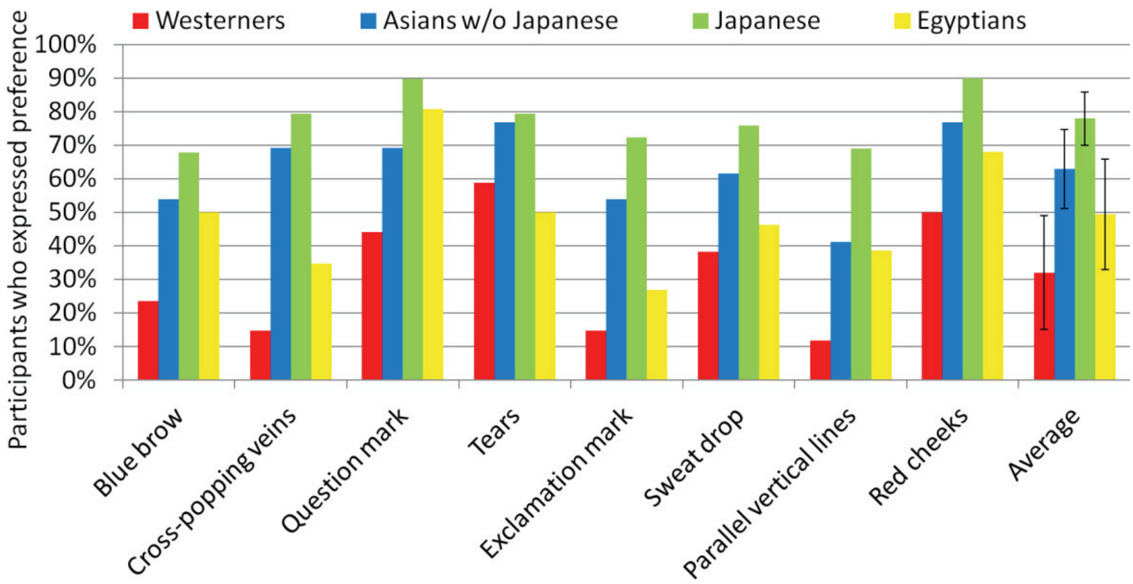


Figure 4.10 Breakdown of preference for the use of symbols by cultural group.

4.4 Context-based evaluation

4.4.1 Introduction

In this part, the purpose was to prove that KOBIAN-R can show human-like capabilities of non-verbal expression. In fact, non-verbal behaviour gives a tone to the whole communication, and sometimes can override the verbal part [147]. For such purpose, written sentences have been used in case of congruence and incongruence with facial expression, and impression on the human partner and degree of credibility of robot's words have been measured.

This section is divided into three different experiments:

i. Evaluation of basic facial expressions with and without congruent and incongruent sentences (Paragraph 4.4.2).
ii. Evaluation of degree of good/bad impression on the human partner and degree of credibility (if robot's words can be believed) depending on the paired face (Paragraph 4.4.3).
iii. Evaluation of impression conveyed by neutral or ambiguous sentences, shown with or without face (Paragraph 4.4.4).

4.4.2 Evaluation of basic expressions and sentences

a. Participants

This survey was done with the collaboration of 42 volunteers (25 male; 17 female; 28.50 average age; 7.89 age standard deviation). Nationality distribution was: 21 Japanese; 8 Italians; 6 Americans; 5 from other European countries; 2 from other Asian countries, for a total of 23 Asian participants and 19 Western participants.

b. Experiment Setup

In order to confirm influence on recognitions of congruent and incongruent face-sentence pairs, a setup similar to Knudsen's experiment [54] was chosen.

First, participants should recognize four basic facial expression, assigning a label choosing between Happiness; Sadness; Fear; Anger. Then, four written sentences were evaluated in the same way, assigning one label each. Finally, faces and sentences were paired in congruent and incongruent way: in case of Happiness and Sadness, the following matches were evaluated again: happy sentence with happy face; sad sentence with happy face; happy sentence with sad face; sad sentence with sad face. The same was done for Fear-Anger. This setup allowed direct comparison with human face recognition made by Knudsen.

Participants had to choose one the answer as in Figure 4.11. The four sentences were:

i. “I can’t tolerate such behaviour"
ii. “Oh no! It’s too dangerous"
iii. “Unfortunately there’s nothing I can do"
iv. “I’m glad things will get better from now on".

Figure 4.11 Example of interface with four choice and incongruent pair.

c. Results

As it can be seen in Figure 4.12 and Figure 4.13, congruence and incongruence seem to have a strong effect on participants' perception. In all the four cases, congruent sentences make expression recognition more clear. The case of Fear deserves to be mentioned, because the facial expression was misinterpreted as Sadness (52.4%). This may be due also to the fact that being the first expression to be evaluated, there could be some bias. Therefore, in future, this kind of tests will be done randomizing or repeating the questions. However, the low score for Fear (45.2%) makes the effect of contextual sentence more evident: matched with a fearful text, correct recognition rate raised to 83.3%.

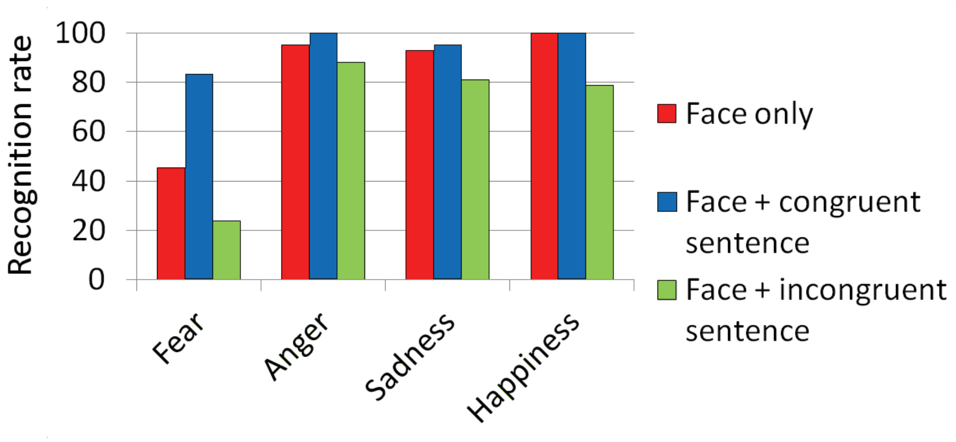


Figure 4.12 Effect of combining two sentences with a facial expression. Recognition rate varies accordingly to congruence.

On the other hand, incongruence has the effect of confusing participants' perception and thus lowering recognition rates. This fact confirms Knudsen's studies and proves that KOBIAN-R is perceived in the same way as humans. Incongruence seems to affect more recognition of sentences (average rate drops by 85.1%) than recognition of faces (average rate drops by 15.5%). This fact is confirmed examining paired emotions in Figure 4.14: excluding Fear, which was misinterpreted, prevalence of face is clear and can be related to prevalence of nonverbal channel on verbal channel, stated by Mehrabian.

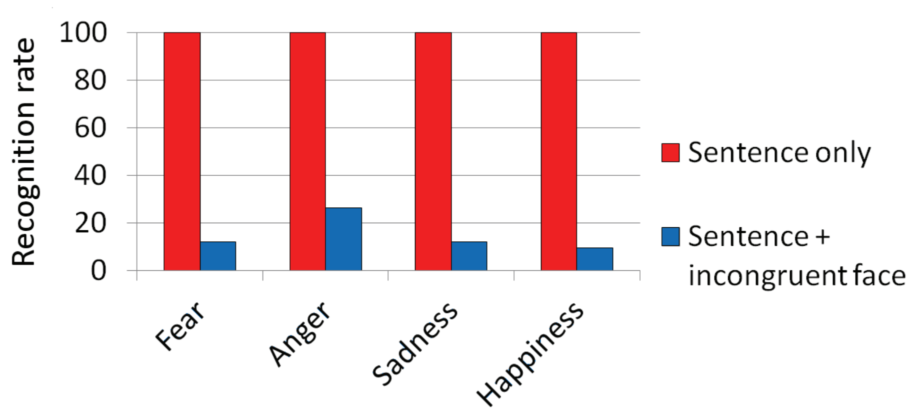


Figure 4.13 Effect of adding incongruent face to a sentence. Recognition rate of the emotion conveyed by the sentence drops significantly in all cases.

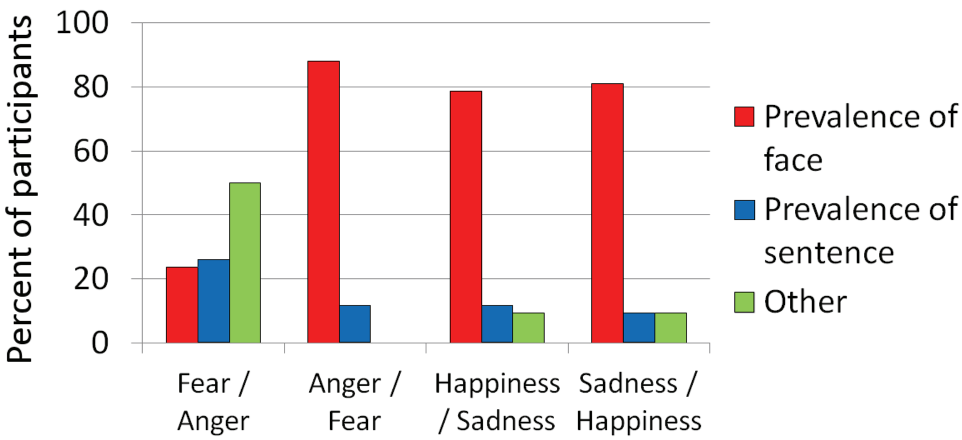


Figure 4.14 Case of mismatch of face and sentence. Facial expression seems to prevail on sentence meaning.

4.4.3 Evaluation of impression and credibility

a. Participants

The participants of this experiment were the same as the previous one.

b. Experiment Setup

The objective of this part was evaluation of Impression (good or bad feeling) and Credibility (how much robot's words can be believed) conveyed by the facial expression-sentence pair, through 5-points semantic differential scales. Two different expressions (one good, one bad, but both appropriate to the context) were shown to the participants together with the sentence. The pairs are shown in Figure 4.15.

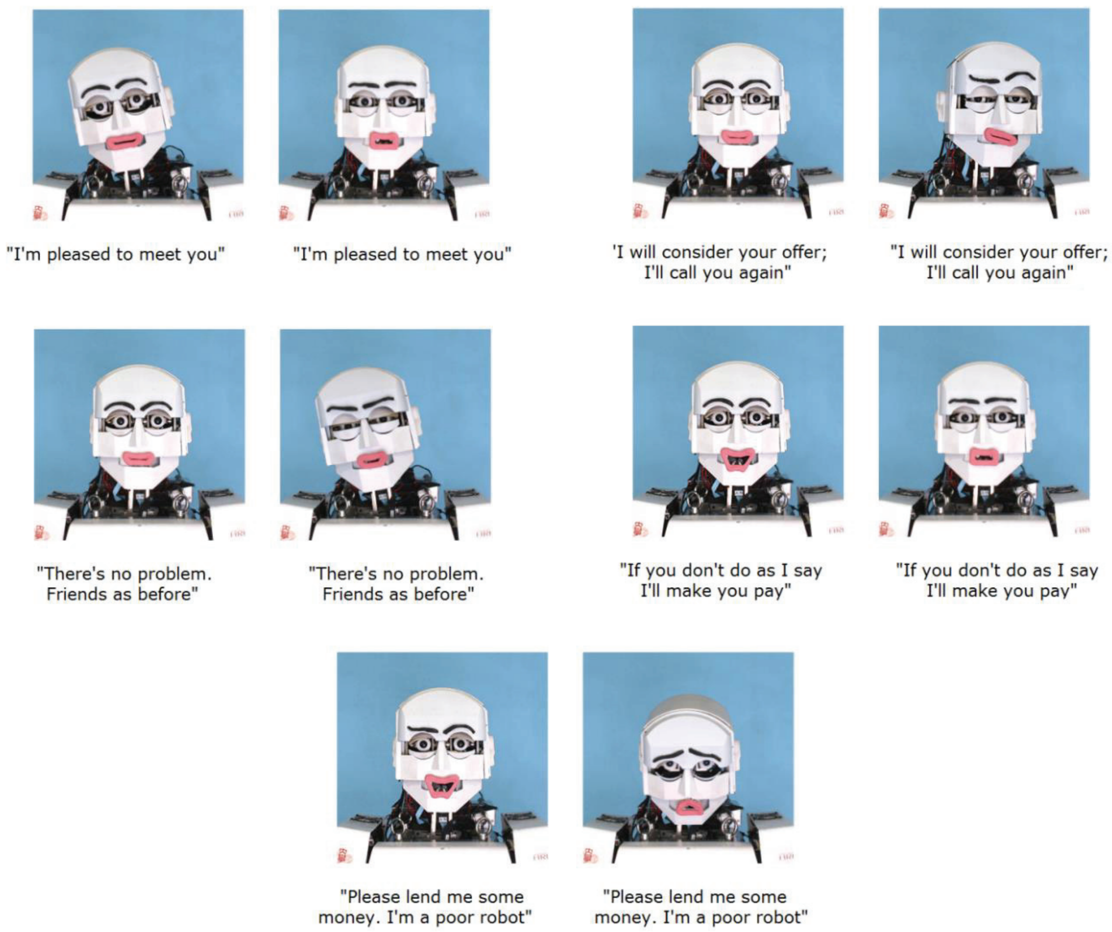


Figure 4.15 The five pairs of facial expressions matched with a sentence. When there is incongruence, the robot results less believable.

Out of the five cases, in the three cases the chosen “good expressions" also implied credibility; in the other two cases, the “bad expressions" were supposed to be more credible, as the written message was negative.

c. Results

In this part of the test, I expected to see in the first three cases good scores for good faces and bad scores for bad faces, for both Impression and Credibility, as the conveyed message was positive. In the latter two cases, where the message is negative, I expected to see the bad face being considered more credible, as a good face would mean joke. The graph in Figure 4.16 confirms these assumptions. Therefore, we can state that congruence makes the robot appear believable, an important attribute that has been discussed in the introduction.

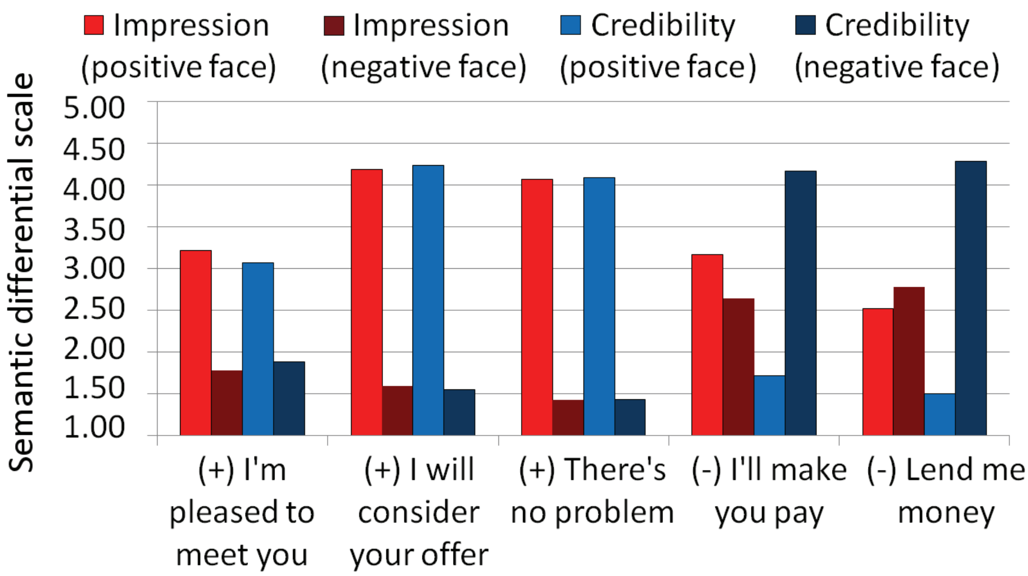


Figure 4.16 Impression and Credibility conveyed by the robot.
They are coherent, respectively, with positiveness of the face and of the sentence meaning.

4.4.4 Evaluation of impression of ambiguous sentences

a. Participants

The participants of this experiment were the same as the previous one.

b. Experiment Setup

The last part featured evaluation of Impression conveyed by five neutral or ambiguous sentences, shown without face as in Figure 4.17.

i. “Oh, you again"

ii. “It's better if you do that as soon as possible"

iii. “Why are you dressed like that?"

iv. “Did you already know that news about my brother Wabian?"

v. “Oh really? How interesting".

Subsequently, evaluation was done again with a couple of expressions which could alter the meaning of the message, as in Figure 4.18. There was no congruent or incongruent pair: it was up to the participant to indicate whether it is perceived good, neutral or bad.

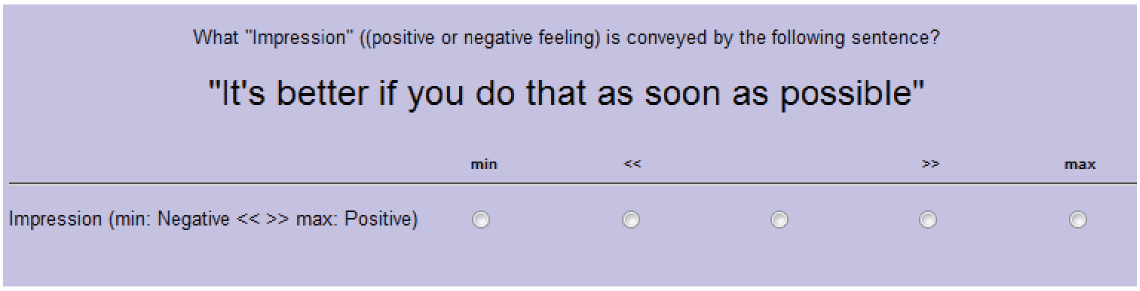


Figure 4.17 Example of evaluation of Impression of a sentence.

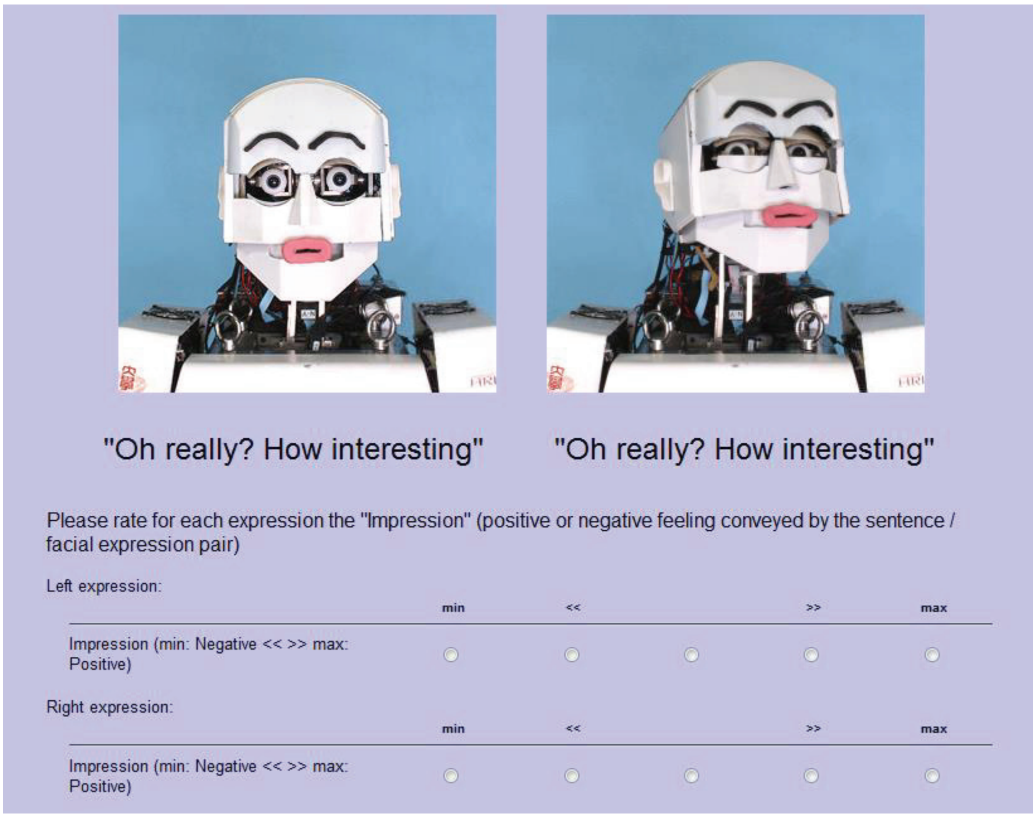


Figure 4.18 Example of interface of the survey, with semantic differential scales.

c. Results

In this last part, sentences were either ambiguous or neutral, and the participant indicated whether they are perceived good, neutral or bad (before and after the addition of faces). Once again, the results (Figure 4.19) confirm my expectation: the addition of faces alter significantly the perception ($p<.05$ for positive faces; $p<.01$ for negative faces). The last case (v) seems different for the others, as it shows a slightly lower Impression rate after the addition of the positive face. This is actually due to the choice of the expression, which was decided to be “Unexpectation", in order to be appropriate to the context, but cannot be considered positive or negative.

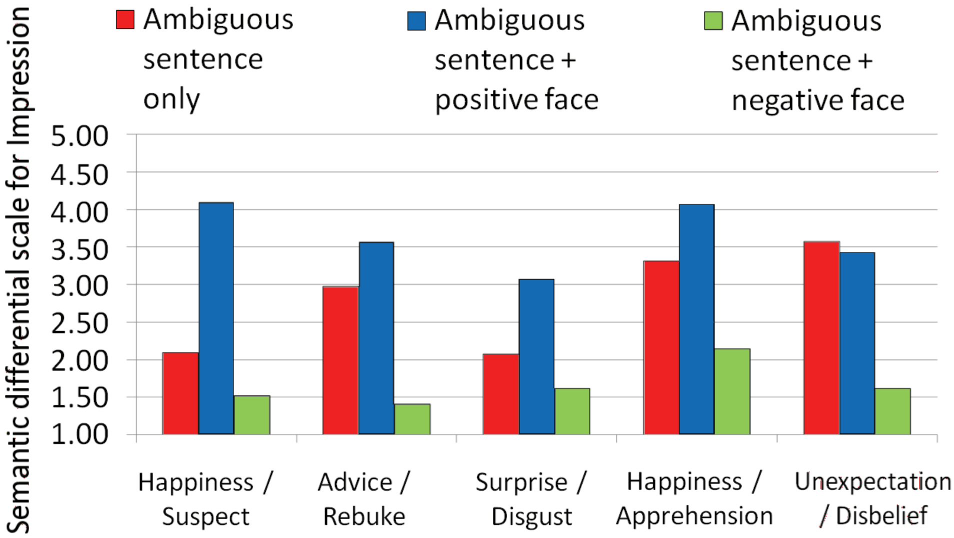


Figure 4.19 Case of mismatch of facial expression and sentence meaning. Impression from facial expression seems generally to prevail.

4.5 Discussion

Given the studies carried out in this chapter, there are some key points that need to be highlighted.

4.5.1 Effect of facial asymmetry

In the study about asymmetry, it was shown that there exists a marked preference for asymmetrical expression in case of negative valence and for symmetrical expression in case of happiness.

Most importantly, the rules extracted from artists' works were not completely confirmed, as the meaning of asymmetrical mouth was misjudged. This fact might be due to the current limitations of the hardware of the mouth. It is therefore necessary to think again about this issue. In this regards, new tests could be done in the future to determine absolute values of recognition rate, in order to pursue the objective of refining the making of asymmetrical expressions.

4.5.2 Evaluation of culture dependent expressions

Results of the study in the previous chapter confirmed the existence of a gap in recognition. For example, the clear result of recognition of "Incomprehension" facial expression by Japanese participants, shown in Figure 4.8, supports the hypothesis that exaggeration of facial cues can help making recognition easier regardless of being less realistic. Therefore, culture-based customisation of facial expressions seems to be effective.

It is my conviction that pursuing this way of expressing emotions may lead in the future to very high scores of recognition rate and thus to a better acceptance and interaction and with humans.

4.5.3 On the use of comic symbols

After displaying comic symbols, results comprehensive of data gathered from Egyptian participants (Table 4.6 and Figure 4.10) show a clear bias between the groups: a total of 66.8% of Western participants do not feel the need of using symbols, compared to 20.7% of Japanese. Moreover, the most important fact we understand from this data is that Westerners and Egyptians recognise as useful only red cheeks, tears and (only in the case of Egyptians) question mark: only such symbols were preferred by at least 50% of the participants. People from other Asian countries show preferences similar to Japanese, who expressed their acceptance for all the symbols. We conclude that these symbols are an important key to dramatically improve recognition.

It is also possible that high-low context differences in Hall's dimensions of culture [148] may play a role in expression recognition. This comes in accordance with findings [149] of cultural differences in people's engagement with a social robot, depending on whether the culture is high context (communication is influenced more by context information and body language) or low context (communication depends highly on direct information).

Using these symbols, the resulting robotic head will appear less realistic, but at the same time it will be able to communicate its emotions more clearly. Some may argue that humanoid robots are supposed to be as similar as possible to humans, as we humans do not need additional "tools" for expressing emotions. However, humans can take advantage of their knowledge of the context, and may actually use alternative ways of expressing emotions when communication channels are limited. For instance, when sending written messages on a chat, emoticons are frequently used. KOBIAN-R's face is limited in some aspects compared to a human face, but on the other hand, it can display exaggerated facial cues. Through the implementation of hardware for symbols display, the robot will benefit of one additional tool to compensate its handicaps. Even if the result will be less human-like, I believe that, like a painting can convey emotions to the same extent and even stronger than a photograph, the best way for the robot to communicate its emotions easily is using additional channels.

As discussed in [57], computer emoticons are perceived differently depending on culture: the current research contributes that each Japanese comic symbol has a different degree of usefulness depending on culture. For this reason, I believe that further research in this direction will lead to a better accuracy in recognition of facial expressions for certain cultural groups may happen, and ultimately to better acceptance of humanoid and social robots.

4.5.4 Context based evaluation

The results of the experiments on context support the idea that KOBIAN-R can show human-like capabilities of nonverbal expression. In fact, evaluation of sentences, together with both congruent and incongruent facial expressions also proved that KOBIAN-R's nonverbal communication influences the overall meaning in the same way as we could expect from human nonverbal cues.

Since expressions seem to be more important than written sentences, the current limits in facial expression recognition, partially caused by cultural bias, are even more an important issue that needs to be addressed. We believe that the use of full body gestures and culture specific expressions will hopefully remedy those problems.

4.6 Summary

In the present chapter, three main studies on facial expressions were performed using photos of the robot KOBIAN-R.

4.6.1 Effect of facial asymmetry

The first study was about asymmetry in the robotic face. I reported results of the investigation of preference of participants for asymmetrical or symmetrical expressions, and the evaluation of influence of asymmetry on a happy face. Results shown that there is a marked preference for asymmetrical expression in case of negative valence and for symmetrical expression in case of happiness. Furthermore, it was my interest to investigate in the present study the effect of culture difference on facial asymmetry perception. From these results, it can be stated that the preferences above mentioned are true for all the cultural groups: data do not prove the existence of any correlation between perception of facial asymmetry and culture.

4.6.2 Cultural study

The second study was focused on culture. The goal of this study was the assessment of culture differences in recognition of a humanoid robot's facial expressions, and finding solutions, as a recognition gap was found. The generator of facial expressions for KOBIAN-R was extended in order to be able to produce different results for different cultures (Western and Japanese), by switching its training set. Such data is based on the work of Japanese illustrators and cartoonists.

The tests were done though web surveys, using photos of expressions in different versions, These tests were repeated with Egyptian participants, who are neither "Westerner" not Japanese. Results confirmed the existence of a gap in recognition and of a bias in preferences depending on cultural group. Relatively low recognition rates for Egyptians pointed out the need to address this problem for Middle Eastern culture.

A set of Japanese comic-style symbols then was defined and those symbols were added to the facial expressions to evaluate its potential use. Results showed that the different perception of written symbols, which are preferred by the majority of the Japanese, but only seldom liked by Westerners or Egyptians, opens new possibilities for compensating the recognition biases. The hardware to display these symbols should be developed. It will be discussed in the future works.

4.6.3 Context based evaluation

The last study, regarding context, led to the following conclusions. As in human face, even in KOBIAN-R's face non-verbal channel seems to be more important than verbal. Impression and credibility of a message conveyed by the robot are coherent, respectively, with positiveness of the face and of the sentence meaning. The addition of face alters the perception of impression significantly. All these features of non-verbal communication of KOBIAN-R are similar to humans, proving that the robot can have human-like non-verbal abilities.

Chapter 5

Greeting interaction

5.1 Introduction

5.1.1 Objectives of this experiment

This chapter contains the description of two cross-cultural experiments in which the reactions of human subjects involved in a simulated video conference with the humanoid robot KOBIAN were observed. The subjects were either Japanese or Egyptian and the robot was greeting and speaking either like a Japanese or an Arab. I expected that Egyptians would have preferred the Arabic version of the robot, and that they might have felt symptoms of discomfort when interacting with the Japanese version. On the other hand, I expected Japanese subjects' perception to be the opposite, i.e. preference for the Japanese version of the robot and discomfort for the non-Japanese one.

We performed the first session of the experiment in Egypt, thank to the collaboration with Egypt-Japan University of Science and Technology, gathering

Egyptian subjects as well as a few Japanese living there. Partial results were published in [150].

We then performed a second session of experiments in Japan using the same experimental protocol. We can now have a look at the complete data (extracted from 61 subjects in total) and compare the cultural groups. The complete work has been published in [151].

5.2 Experimental setup

5.2.1 Hardware

In order to make an experiment with subjects in a place like Egypt, distant from the robot (which is in Waseda University, in Tokyo, Japan) a video conference system is needed. Despite there is only one KOBIAN, our purpose was to show two different robots (one Japanese-like, one Middle Eastern-like) to the subjects; therefore, the video conference was simulated. We used the robot in two versions: KOBIAN, the original version, and AL-BIAN, which has different facial and body colours (see Figure 5.1, a and b respectively). The colour differences between the two versions were chosen to be not related in any way to the specific culture, and they are not meant to make the robot more appealing for a specific group of subjects; their only purpose is to give to the subjects the impression that they are interacting with two different, although very similar, robots.

KOBIAN and AL-BIAN were used to realise the culture-specific greetings (motion of the arms and waist) and to simulate speech (motion of the lips and slight periodic oscillations of the head, that give a human-like appearance to the robot behaviour).

The robot body parts are controlled by both position-based and velocity-based controllers that have been implemented using YARP [152], a software framework for robot programming. The coordination of the different joints involved in the motion and the timing of the different movements were accurately designed to achieve a natural behaviour with smooth trajectories and mild transitions between the different motions.

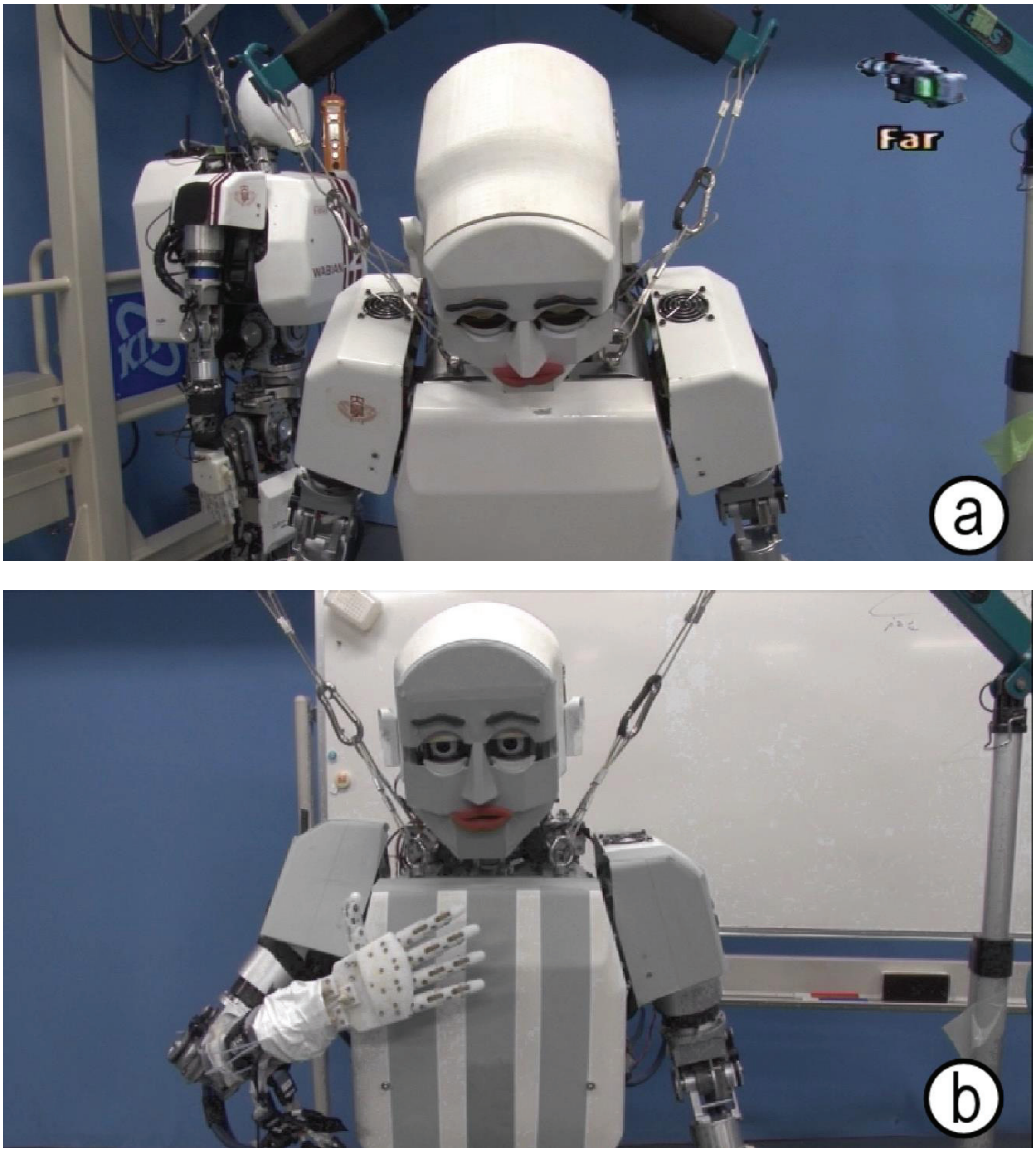


Figure 5.1 Two screenshots of the video shown to the subjects.
(a): the robot KOBIAN performing a bow. (b): the ficticial robot AL-BIAN performing a salaam greeting. They are in fact the same robot.

5.2.2 Experimental protocol

The experimental protocol consists of the following 8 steps:

Step 1. Pre-questionnaire

Each subject is invited to sit at a desk, in front of a big screen, and to compile a preliminary questionnaire on likeability of humanoid robots in general and on their own perceived safety (details in Paragraph 5.2.4).

Step 2. Explanation

The subject (Figure 5.2, a) is explained the purpose of the experiment and he/she is told there will be a call to a laboratory in Waseda University in Japan through the video conference system, for showing two different robots. Actually, a previously recorded video will be shown, as the TV is not connected to the device, but to a PC. No actual call is made, but the subject is tricked into believing that he/she is watching a live connection by adding the typical connection sounds and screenshots. This Wizard of OZ style experimental setup encourages natural behaviour of the subject.

Step 3. Examiners preparation

One of the two examiners (Figure 5.2, b) is in charge of measuring the response time of the subject's greeting (either spoken or a gesture) to the robot's greetings by using a stopwatch. The other examiner (Figure 5.2, c), who controls the PC (Figure 5.2, d) and the video conference system remote control, sits in front of the subject for examining any verbal or non-verbal cue expressed by the subject; he also takes notes using a checklist.

Step 4. First call

As one examiner pretends to start the call, video begins and connection is established with a Japanese student, who once more explains the purpose of the experiment; then the Japanese student switches the camera to

KOBIAN, who greets, does a self-introduction and says goodbye (more detail in Paragraph 5.2.3).

Step 5. First questionnaire

After closing the connection, the subject compiles a questionnaire about KOBIAN, including all the questions shown in Paragraph 5.2.4 .

Step 6. Second call

A new call is made, this time to an Arabic speaking student who greets and tells the subject to wait, then switches the camera to AL-BIAN, who greets, does a self-introduction and says goodbye.

Step 7. Second questionnaire

As the video conference ends, the subject is invited to compile a questionnaire about AL-BIAN, including all the questions shown in Paragraph 5.2.4 , and to express a preference between the two robots.

Step 8. Closing explanations

At the end, the subject is informed that the video conference was not real, and of the motivation of the use of this trick. If he/she knew beforehand that was watching a video, there would be no reactions, and no interaction. Through this trick, we could collect meaningful data from their spontaneous reactions (in fact, as we later verified, nobody noticed the trick).

Note: for all subjects the order of the robots was randomly chosen (steps 4-7). This means that for around one half of the subjects, the order of the robot, instead of (KOBIAN, AL-BIAN as in Figure 5.3 and Figure 5.4) was (AL-BIAN, KOBIAN).

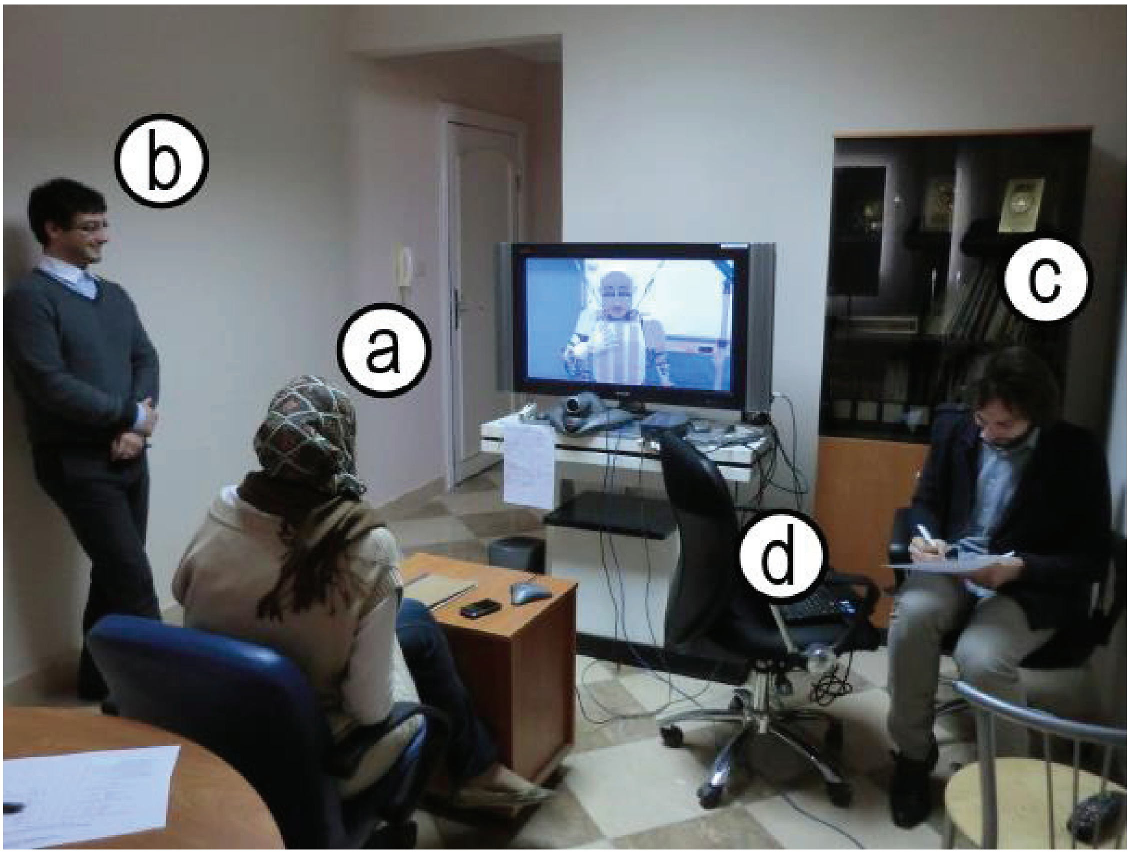


Figure 5.2 Experimental setup during a video call in Egypt.
The subject (a) is watching the screen, and the two examiners (b and c) are focusing on the subject and controlling a computer (d)

5.2.3 Videos

Several videos were recorded beforehand and assembled together into a single video file. Interface screens and sounds were added for simulating a real call through a video conference system. The video, shown in Figure 5.3 and Figure 5.4, was composed by the following parts:

i. Japanese person greeting in Japanese, introducing in English the next robot;
ii. KOBIAN performing a bow – with Konnichi wa (which means Hello or Good day) speech added – as initial greeting;

iii. KOBIAN introducing himself in English with Japanese accent;

iv. KOBIAN performing a bow – with Otsukaresama desu (which is a standard idiomatic phrase that fellow workers use at the end of a working day) speech added – as final greeting;

v. Middle Eastern person greeting in Arabic, and introducing in English the next robot;

vi. AL-BIAN raising hand – with Alsalamo alikum (which means Hello or Good day) speech added – as initial greeting;

vii. AL-BIAN introducing himself in English with Arabic accent;

viii. AL-BIAN moving its hand on the heart and nodding (a shortened version of the Mouth-and-forehead salaam greeting described in [76]) – with Alsalamo alikum speech added – as final greeting.

5.2.4 Assessment

A combination of physiological responses and written questionnaires were considered for the assessment of the interaction, since in this way it is possible to catch both explicit opinions and psychological reactions.

For assessing the degree of discomfort of subjects, a good method is to observe eyebrows frowning, through measurement of movement of the corrugator supercilii muscle [153], [154]. In fact, frowning is known to be a symptom of either incomprehension or anger [155].

However, in our pilot study, facial Electromyography did not provide reliable data; furthermore, electrodes placed on subjects' face caused discomfort to them. Therefore it was decided to use an examiner who would observe facial expressions, non-verbal cues (not limited to frowning) and behaviour of the subjects, and compile a checklist with all these relevant information.

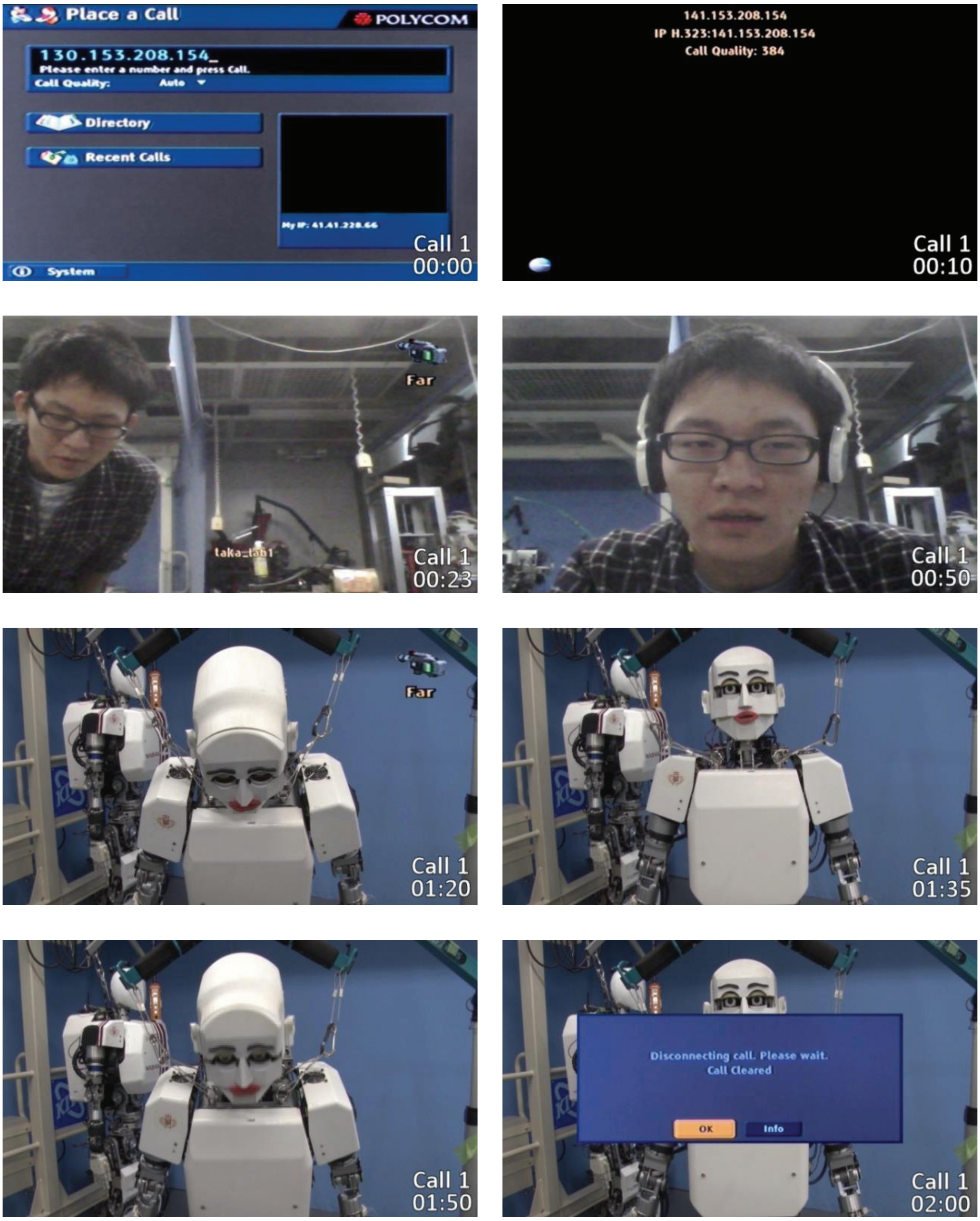


Figure 5.3 Screenshot sequence of the first call.

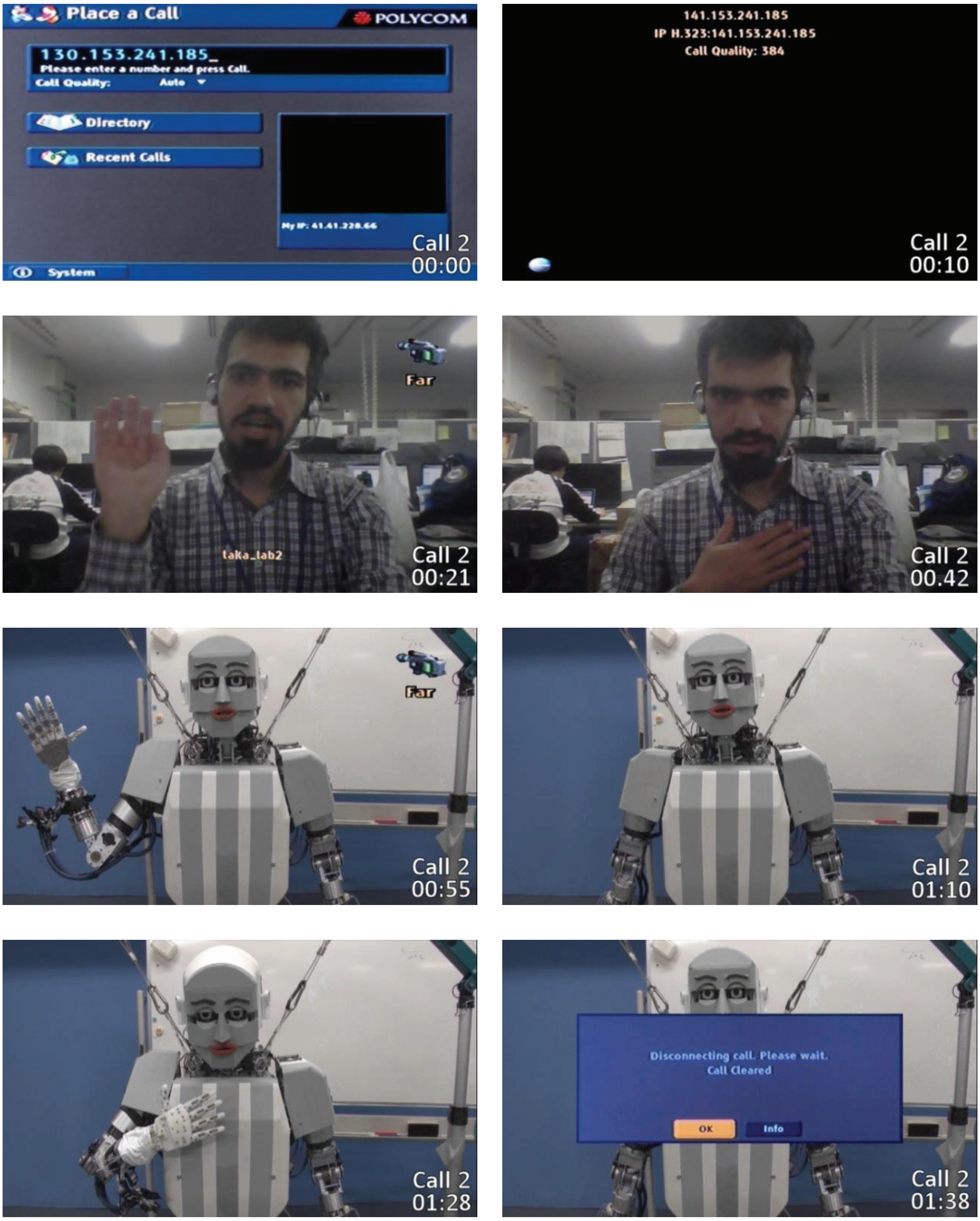


Figure 5.4 Screenshot sequence of the second call.

Assessment in human-robot interaction through survey is preferably done using standardised measurements. Bartneck [156] devised reliable 5-point semantic differential scales called Godspeed for measuring anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety for robots. I decided to use likeability and perceived safety; moreover, I added a new set of scales for measuring cultural closeness. The three resulting groups of scales (first and third from [156]) were presented as follows:

LIKEABILITY						
Dislike	1	2	3	4	5	Like
Unfriendly	1	2	3	4	5	Friendly
Unkind	1	2	3	4	5	Kind
Unpleasant	1	2	3	4	5	Pleasant
Awful	1	2	3	4	5	Nice
CULTURAL CLOSENESS						
Impolite	1	2	3	4	5	Polite
Mysterious	1	2	3	4	5	Familiar
Incomprehensible	1	2	3	4	5	Comprehensible
Foreign	1	2	3	4	5	Native
PERCEIVED SAFETY						
Anxious	1	2	3	4	5	Relaxed
Agitated	1	2	3	4	5	Calm
Quiescent	1	2	3	4	5	Surprised

Additional questions included some demographic information like age and gender, and some more explicit questions regarding what the subject liked about the two robots. Moreover, some specific questions were made about the gesture and the words the robot used, and the way it spoke English. In the analysis of results all the significant answers collected from these questions are shown. Questionnaires were written in two languages (English and Arabic or English and Japanese). The English and Arabic version is shown in Appendix B.

5.2.5 Participants

The whole experiment was done in two sessions. The first one was done in Egypt, inviting 36 subjects; the second session was done in Japan, with 25 subjects. The experimental setup was the same in both locations. In total, we could gather the data of 61 subjects (male: 37; female: 24; average age: 30.33; standard deviation: 10.29). We gathered a heterogeneous group consisting of people with different age and education level, rather than just students. The unbalance between male and female happened because it's not easy to find female subjects in Egypt available to do an experiment, compared to men. As a consequence of this unbalance, an analysis on differences between genders might produce misleading results and therefore it was not carried out. The total of the subjects were instead divided in four groups:

- Group J: Japanese people with no previous experience with Middle Eastern culture (18 subjects);
- Group JE: Japanese people living in Egypt (5 subjects) or with some degree of interest in Arabic language or Middle Eastern culture (7 subjects). (Total: 12 subjects);
- Group EJ: Egyptian people who can speak Japanese, or have been in Japan or have interest in Japanese culture (13 subjects);
- Group E: Egyptian people who have no previous experience with Japanese culture (18 subjects).

As mentioned previously, the order of the two robots ((KOBIAN, AL-BIAN) or (AL-BIAN, KOBIAN)) was randomly chosen for each subject in every group. This manipulation, while useful for avoiding a bias, did not produce any significant effect on the results, therefore it is excluded from further analysis.

The duration of the experiment was approximately 20 minutes.

5.3 Data analysis

Gathered data were analysed using the Kruskal-Wallis test [157] and subsequently the Mann-Whitney U-test [158]. In all the cases in which the U-test was performed, it means that the Kruskal-Wallis test already gave a low *p* value as output. The classical Student's t-test and ANOVA could not be applied, because the shape of the distribution graph resulting from the semantic differential scales data was not a normal distribution. In all the following graphs, one asterisk (*) means $p < .05$; two asterisks (**) mean $p < .01$; three asterisks mean $p < .001$.

5.3.1 Subjects' preference

At the very end of the experiment, the subjects were asked to express their preference between the two robots. Result of this explicit question shown that while Japanese subject of group J have a strong preference for KOBIAN (Figure 5.5, on the left-hand side), Egyptian subjects prefer AL-BIAN (Figure 5.5, on the right-hand side), and the other groups stand in the middle. In Figure 5.5 average values are highlighted when the U-test produced significant results ($p < .05$).

Subjects were also asked to justify their choice adding a free comment. We have collected all the comments and divided into the following categories, shown in Figure 5.6:

a. Non-verbal communication: gesture more natural / uses hands / moves hands like humans / better body language / better movement / more realistic;

b. Sense of familiarity: it is more comfortable / more familiar / more friendly;

c. Language: speaks my language / greeting is in Arabic;

d. Understandability: more understandable / clear language / clear spelling / voice is more clear / speaks more fluently / better English;

e. Emotion: emotion more clear / shows emotions better;

f. No reason: I don't know / just my feeling

I included any comment related to the appearance to the "No reason" category. This is because physical appearance of the two robots was essentially the same, and claiming that one of the two is more good looking may be caused by personal feelings.

Additional negative comments were made about the quality of the voice, but as these comments were made on both robots, we believe that this is not an important factor for our evaluation and therefore did not include it in the diagram.

The categories most important resulted to be the sense of familiarity, understandability and non-verbal communication.

As integration to the explicit preference, we asked one additional question to the subjects: "Would you like to meet this robot again?", using again a 5-point semantic differential scale. Results are coherent with the ones seen so far: subjects of group J would prefer to meet KOBIAN (+0.39 compared to AL-BIAN); Japanese of group JE do not have a strong preference (+0.08 for AL-BIAN). Egyptians would like to meet AL-BIAN (+0.23 for group EJ, +0.44 for group E compared to KOBIAN).

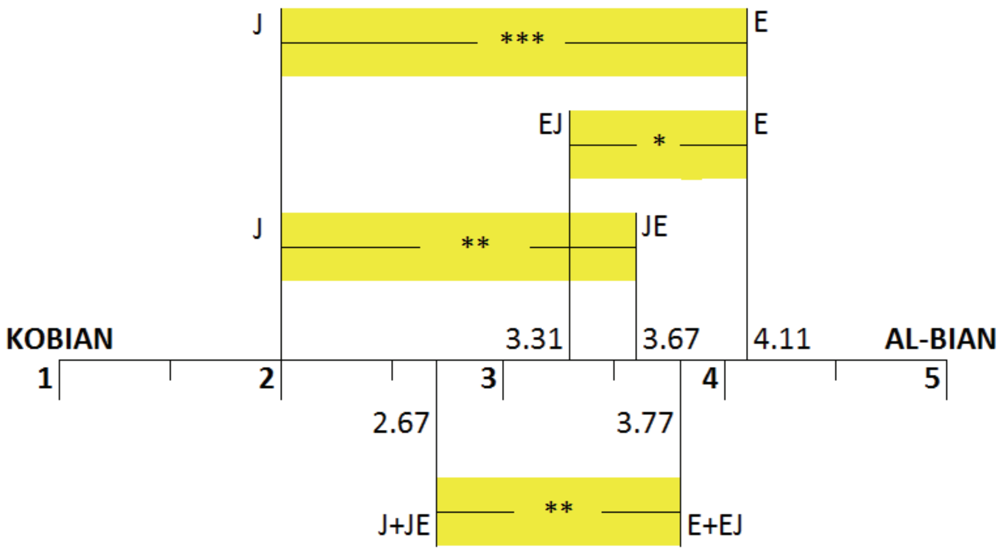


Figure 5.5 Significant difference in preference between KOBIAN and AL-BIAN. The different groups J, E, EJ and JE are compared and between the total of the Japanese (J+JE) against the total of the Egyptians (E+EJ). In yellow, statistically significant differences (one asterisk (*) means $p < .05$; two asterisks (**) mean $p < .01$; three asterisks mean $p < .001$).

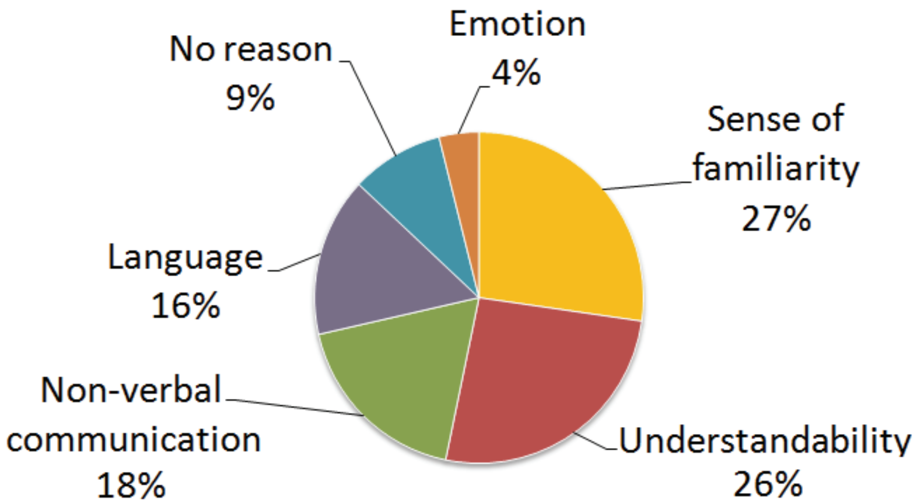


Figure 5.6 Graph of the reasons why subjects expressed a preference for one robot.

5.3.2 Cultural closeness

For measuring cultural closeness, I introduced a new set of scales, described in Paragraph 5.2.4 and presented to subjects in steps 5 and 7 of the experiment protocol. These scales can be considered reliable, as their Cronbach's alpha, a coefficient of internal consistency used to estimate the reliability of a psychometric test, is greater than 0.7 [159]. Subtracting the average score (from 1 to 5) obtained by AL-BIAN and KOBIAN, we can measure perceived cultural distance of the two robots for each group. Results in Table 5.1 show that this difference is significant for the groups E, who feel AL-BIAN closer, and J, who feel KOBIAN closer. In addition, differences among groups are also present. Group JE shows on average almost no preference between the two robots, and is significantly different from Group J ($p < .05$). A much stronger difference ($p < .001$) can be found between groups J and E and between the total of Japanese subjects (J + JE) and the total of Egyptian subjects (E + EJ).

Table 5.1 Difference in cultural closeness of the two robots by group

	AL-BIAN – KOBIAN	Between groups
J	–0.61 (**)	J–JE: * ; J–E: ***
JE	–0.04	
EJ	0.25	
E	0.514 (*)	

In yellow, statistically significant differences between the two robots in a single group, in the cells, and between two groups, on the right. One asterisk (*) means $p < .05$; two asterisks (**) mean $p < .01$; three asterisks mean $p < .001$.

5.3.3 Likeability and perceived safety

Both likeability and perceived safety were measured three times, namely at steps 1, 5 and 7 of the protocol. We consider the measurement of step 1 a preliminary assessment (Pre in Figure 5.7) of the subject's attitude towards humanoid robots in general (not on the two specific robots) and it is useful to give a hint of the acceptance of the robots in absolute terms.

Significant data for likeability are as follows: 3 groups out of 4 (E, EJ, JE) showed a significant preference for AL-BIAN, whereas the score for group J was relatively low. On the other hand, before drawing conclusions on perceived safety, one of the scales (Surprised/Quiescent) has to be dropped, because it does not seem to be consistent with the other scales, leading to a low Cronbach's alpha (< 0.7). As a result, average is calculated on the two variables Anxious/Relaxed and Agitated/Calm. Probably because of this inconvenience, no significant shift of perceived safety before/after meeting the robots is detected. Kruskal-Wallis test also confirms this assumption (p = 0.23). Nevertheless a trend can be noticed in Table 5.2: meeting AL-BIAN seems to have a slightly better effect for all groups except group J.

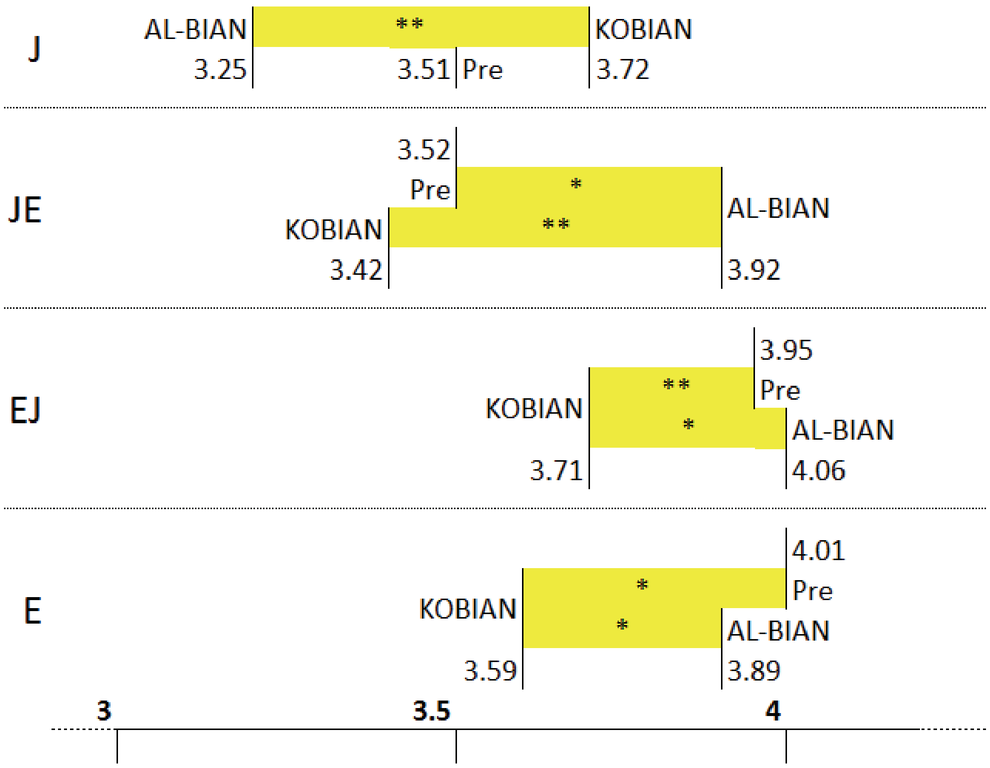


Figure 5.7 Zoom in variations of Likeability on 1 to 5 scale.
In yellow, statistically significant differences: one asterisk (*) means $p < .05$; two asterisks (**) mean $p < .01$; three asterisks mean $p < .001$. "Pre" indicates the preliminary assessment on Humanoid robots done before doing the experiment.

Table 5.2 Differences in perceived safety before and after interacting with the robots

	KOBIAN – Pre	AL-BIAN – Pre	Difference
J	0.14	−0.25	KOBIAN: +0.39
JE	−0.21	0	AL-BIAN: +0.21
EJ	0.23	0.42	AL-BIAN: +0.19
E	−0.19	0.14	AL-BIAN: +0.33

In yellow, negative values: the subject feels more anxious afterwards; in green, positive values: the subject feels more safe. In lime: no difference. "Pre" indicates the preliminary assessment on Humanoid robots done before doing the experiment.

5.3.4 Reaction to greeting types

We recorded detailed information about each subject's reaction to the greetings of both human operators (we call here human operators the two students who introduced the robots during the video conferences: see Paragraph 5.2.2 , steps 4 and 6) and robots. Particular emphasis was placed on measuring response time of the subject after human/robot greeting speech, since a delay could be a hint of hesitation and feeling of bewilderment. From Table 5.3, comparing all matches, it can be seen that subjects experienced hesitation mainly with robots rather than with humans. For Egyptians, the highest scores (considering long reaction time as > 1 second) were hit when facing KOBIAN, while it is not possible to judge Japanese subjects, as the interaction with AL-BIAN was too poor. In general, it is clear that the amount of interaction with the two robots is still low compared to humans (rows labelled Answer in Table 5.3, where the answer can be either spoken or by gesture).

Table 5.3 Percentage of cases of answer and of long reaction time

		Human (Japanese)	**KOBIAN**	**Human (Arabic)**	**AL-BIAN**
J	Answer	75%	20.8%	18.1%	4.2%
	Slow response	0%	14.6%	0%	0%
JE	Answer	62.5%	18.8%	27.1%	0%
	Slow response	0%	0%	0%	0%
EJ	Answer	61.54%	26.92%	50%	7.69%
	Slow response	0%	29.17%	5%	0%
E	Answer	38.9%	12.5%	61.1%	26.3%
	Slow response	0%	17.5%	4.51%	9.38%

5.3.5 Relative weight of speech and gestures

In order to understand which factors, among gestures and voice, had a stronger impact on subjects' preferences, the questionnaire featured the following questions (to be answered with the 5-points semantic differential scale):

- Do you like the gesture the robot used to greet you?
- Do you like the words the robot used to greet you?
- Do you like the way the robot speaks English?

As a result, we got the data displayed in Table 5.4. It appears that both channels of communication contribute to the preference.

The same investigation was extended to compare robots' greetings with the human operators' greetings. In fact, the two operators performed exactly the same greetings, in terms of speech ("Konnichi wa" and "Alsalamo alikum") and gestures, of the two robots; thus they can be compared.

Given some comments made by the subjects, we hypothesised that the ones who belong to a certain culture are stricter when evaluating an imitation of their own way of greeting. For example, a Japanese is more likely to notice any incorrectness in the angle of the bow. We expected this kind of bias to be stronger in groups J and E.

Table 5.4 Differences in preference of greeting due to verbal and non-verbal channels

	Gesture	Greeting words	English speech
J	KOBIAN +0.56	KOBIAN +0.72	KOBIAN +0.11
JE	AL-BIAN +0.25	AL-BIAN +0.17	AL-BIAN +0.42
EJ	AL-BIAN +0.22	AL-BIAN +0.08	AL-BIAN +0.77
E	AL-BIAN +0.17	AL-BIAN +0.56	AL-BIAN +1.39

However, this hypothesis could not be verified through the questionnaire results and no significant conclusion can be drawn. For the sake of completeness, we report here these results in Table 5.5. Row data contain no significant high or low (average is 3.82; standard deviation is 0.31). Some small differences can still be noticed. For example group J prefers the robot in case of Japanese greetings, but the human's in case of Arabic.

Table 5.5 Differences in preference of greeting compared to human operators

	Gesture (Japanese)	**Speech (Japanese)**	**Gesture (Arabic)**	**Speech (Arabic)**
J	KOBIAN +0.44	KOBIAN +0.22	Arabic operator +0.17	Arabic operator +0.22
JE	KOBIAN +0.50	KOBIAN +0.42	KOBIAN +0.17	AL-BIAN +0.33
EJ	KOBIAN +0.23	KOBIAN +0.08	Arabic operator +0.31	Arabic operator +0.38
E	Japanese operator +0.11	Japanese operator +0.17	Arabic operator +0.44	AL-BIAN +0.06

5.3.6 Non-verbal communication

One additional proof of subjects' feeling can be obtained by analysing non-verbal cues. Positive ones include smiling, laughing and nodding. Negative ones include eyebrow frowning, eyelids tightening, head shake and similar neck movements.

In Table 5.6 it is possible to see how negative cues are concentrated in the interaction between KOBIAN and Egyptians of group E, and between AL-BIAN and Japanese of group J, proving the discomfort experienced by subjects, and some particular examples are shown in Figure 5.8 and Figure 5.9.

Table 5.6 Average number of occurrences of non-verbal cues per subject

		Human (Japanese)	KOBIAN	Human (Arabic)	AL-BIAN
J	Negative	0.06	0.33	0.11	0.78
	Neutral	0.00	0.11	0.39	0.56
	Positive	0.17	0.72	0.39	0.33
JE	Negative	0.00	0.17	0.08	0.08
	Neutral	0.08	0.00	0.08	0.08
	Positive	0.00	0.50	0.25	0.58
EJ	Negative	0.00	0.31	0.00	0.00
	Neutral	0.00	0.08	0.00	0.08
	Positive	0.15	0.62	0.38	0.54
E	Negative	0.11	0.83	0.06	0.06
	Neutral	0.00	0.17	0.06	0.06
	Positive	0.28	0.50	0.39	0.56

In red, values > 0.75; in orange, values > 0.5; in yellow, values > 0.25.

5.4 Discussion

In these experiments, my goal was to investigate the different attitudes of Egyptian and Japanese people towards two different versions of the same robot, programmed to greet and speak with Egyptian-like and Japanese-like manners. I expected the reactions of the subjects, both explicit and implicit, to be different according to their culture. Data were collected in different modalities: using questionnaires for investigating both explicit comments and implicit effects on subjects' emotional state, measuring response of interaction, and keeping track of all verbal and non-verbal cues.

Results confirmed our hypothesis that Japanese subjects and Egyptian subjects perceive the two humanoid robots differently. Japanese seem to prefer KOBIAN, whereas Egyptians seem to prefer AL-BIAN. This can be seen in Paragraphs 5.3.2 and 5.3.3 in terms of attitude towards the robots, and in Paragraphs 5.3.4 and 5.3.6 , in terms of interaction. Some interesting points, explained in the next paragraph, can be deducted by combining all the data.

Figure 5.8 Example of some non-verbal cues for Egyptian subjects.
Negative cues on the left, and positive ones on the right.

5.4.1 Different impressions of the same robot

Egyptians feel in some cases discomfort when interacting with KOBIAN (paragraphs 5.3.4 and 5.3.6), and even end up in a more anxious state compared to before the experiment (Paragraph 5.3.3). AL-BIAN, who (citing a subject's comment) "Moves hands like humans" does not seem to cause such anxiety. On the other hand, a few comments we gathered from Japanese subjects explain their point of view on AL-BIAN: "Ayashii" (suspicious), "Iwakan" (discomfort), "Tsumetai" (cold), "Kowai" (scary). Such words were not used when commenting KOBIAN, and it is odd, because the appearance of the two robots is almost the same.

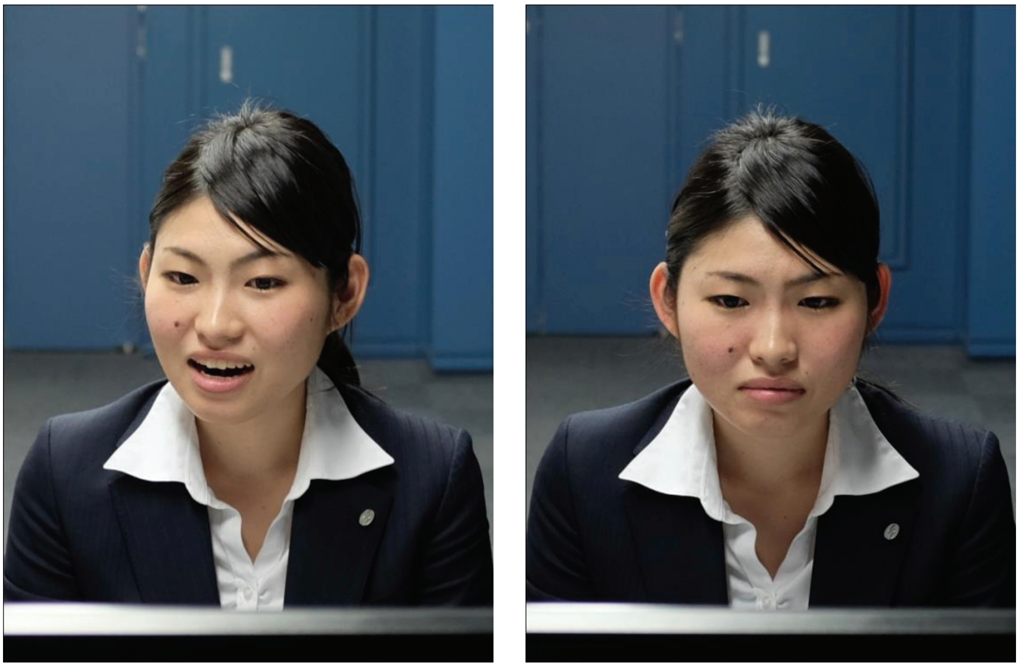


Figure 5.9 Example of neutral and negative non-verbal cues for a Japanese subject. We consider astonishment (left image) as neither positive nor negative.

Subjects familiar with the other culture tend to like both robots and react in the same way. The tendency of groups EJ and JE to stand in the middle of the other two groups in terms of results, is common to most of the gathered data, including subjects' preference and cultural closeness (paragraphs 5.3.1 and 5.3.2). Among these group, the presence of subjects not only familiar, but also enthusiastic about (in case of Egyptians) Japan or (in case of Japanese) Middle East, might explain the cases in which attitude of Egyptians towards KOBIAN is even more positive that AL-BIAN (see Table 5.3) and in which Japanese prefer AL-BIAN (see Table 5.2, Figure 5.5 and Figure 5.7).

It is possible to notice from the data many hints that Egyptians not familiar with Japanese culture have a strong preference for AL-BIAN and that Japanese not familiar with Middle East have a strong preference for KOBIAN. The analysis of the reasons of this preference reveal that causal relationship among the different categories of reasons (shown in Figure 5.6) might exist. In particular, it is possible that gesture and words play a role in making the robot more familiar. In this regard, interaction mechanisms described in The Media Equation [160] should be considered, as similarity to a computer agent [161] and politeness [162] are proved to play a role in human perception of a machine.

An open question is whether roboticists should really need to take care about even small cultural differences. I believe that in the future, when robot might enter the mass market, these small details can make a difference between a product that gets sold and another that does not get sold. Designers might be interested in knowing what to think about when designing for a robot in terms of appearance as well as behaviour. In this experiment, I made AL-BIAN look very similar to KOBIAN; however, the use for instance of clothes (in a similar way to Ibn Sina [21]) might make significant differences between groups bigger.

5.4.2 The impact of language

Is it just a matter of language? Indeed, it could be argued that the results we obtained in this experiment are somehow obvious because of language barrier. However, a closer look to the numbers supports our belief that this is not the case. For example, I can cite the evidence of a Japanese subject, completely unfamiliar with robots and not very interested in the experiment. This subject's answers were '3' for all the question in the 5-point semantic differential scales, for both robots. In spite of this pronounced non-preference, she chose KOBIAN in the last question, with no special reason. This fact suggests that there are some subtle factors which influence people's judgement other than language. In fact, in Figure 5.6 the categories Understandability and Language, put together, correspond to 42% of the reasons of preference. The rest is due to other reasons. In order to shed more light on this matter, there is the need of a further experiment in which the two robots will both speak a language that is not the subjects' mother tongue, whereas the non-verbal part will be, for each robot, respectively belonging to a culture that is considered close to the subjects' one, and to a culture that is considered distant. Another possible test would be inverting the two communication channels: Japanese verbal content with Egyptian non-verbal, and vice versa.

5.4.3 Encouraging interaction

As seen in Table 5.3, amount of interaction with the two robots is still low compared to humans (see Figure 5.10 as one case of successful interaction with a robot). For this reasons, we further investigated and asked to the subjects the reason why they did not reply to the robot but they did to the human operators (who were recorded as well). The most common answers were:

- "I didn't think the robot would listen to me"
- "I thought it wouldn't answer"
- "It made me agitated"
- "I was too shy to answer the robot"
- "I don't think I will be considered impolite if I don't answer to the robot"
- "Not feeling its presence"
- "Not enough eye contact"

The comment about politeness is particularly interesting, because in order to make humans and robots, in the future, develop a more natural interaction, culture and politeness have to be considered. In fact, as hypothesised in a formula made by Brown and Levinson [92], politeness is dependent on culture.

Taking a further look at Table 5.3, it is possible to notice that the interaction gap reduces when robot's greeting matches language and gesture of subject's background culture. This fact suggests that interaction can be encouraged by a robot that looks more familiar and it supports the need of developing culture-specific customisation of robots.

Differences in national culture, history, and religion are known to have an impact on the design of products [163]. Even in software, when people from different cultures look at an object in an interface, they may have a different understanding of what it represents [164]. As robots need to interface with humans, I believe that the concept of localisation of products needs to be extended to robots too, and the results here reported seem to support this hypothesis.

Figure 5.10 Successful moment of greeting interaction.
Even for Japanese subjects, this happened seldom.

5.5 Summary

In the present chapters, two experiments of Human-Robot Interaction, performed in Egypt and Japan, are described. Subjects of the two nationalities participated to a simulated video conference with two robots which performed greetings and a self introduction, respectively using Arabic and Japanese gestures and way of speaking. The purpose of this work was to prove that a robot that can adapt to the verbal and non-verbal communication styles of a specific culture may make a better impression and reduce discomfort of human subjects.

Results suggested the existence of difference in perception between Egyptians and Japanese, as the former prefer the Arabic version of the robot and experience several symptoms of discomfort when interacting with the Japanese version. The same things happen the other way round. This fact supports my hypothesis; therefore, design of a robot which moves and speaks in a way that is linked to a certain cultural background should be carefully considered when thinking about integration of assistive robots into society.

Results of the present study also highlight the need of diversification of robots and justify the realisation of a system for greeting mode selection for humanoid robots, which will be described in the next chapter. Furthermore, a repetition of the same experiment in a third Western country is planned and will be described in the future works.

Chapter 6

Greeting selection

6.1 Introduction

In the Egyptian-Japanese study, culture-dependent acceptance and discomfort were found. As a result, the importance of culture-specific customisation for acceptance of robots is confirmed: robots need a system of greeting selection.

Through the development of a model of greetings and using multimodal input and output, we will teach a robot how to greet. Robots would be able in the future to switch between different modes depending on the cultural background of the human partner.

We can think, as example, of a typical scenario: a foreigner in a country visited for the first time (e.g. a Westerner in Japan) will greet in inappropriate way as long as he is unaware of the rules that define the greeting choice. For example, he might want to shake hands or hug, and will receive a bow instead. The same is true for the use of greeting words depending on the context, as reported by Mizutani et al [165].

While visit the foreign country, after a few iterations that include failed attempts, the foreigner will learn the appropriate way to greet. In fact, humans know unconsciously the factors that influence the choice of greeting, or they are either able to learn them. Robots don't. Can a robot learn in the same way?

Objectives of this section are:

- Make a novel model of the greetings that a robot can learn and use.
- Make such model flexible, so that a robot can switch between different modes for manners of each culture.

Thanks to the collaboration with the Karlsruhe Institute of Technology, in Karlsruhe, Germany, the experiment was done using the robot ARMAR-IIIb.

6.2 Greeting choice factors

In the introduction I described the state of the art in sociology of the factors that influence the choice of greeting gesture and words. This is the starting point for making a model of greeting selection. Figure 1.4 contained the overview of the factors. It is clear that as it is, the graph is too complex to be usable. It needs to go through a process of simplification: in Figure 6.1, the factors to be cut are greyed out.

The simplification was done following these assumptions:

a. Only 2 individuals (robot and human participant): I prefer to circumscribe the problem to the encounter of only two parties, without taking consideration a higher number of individuals.
b. Eye contact is taken for granted: as the establishment of eye contact is a problem of machine vision, let us suppose that the two parties meet face to face. The experimental setup will also need to be planned carefully to ensure that this happens.
c. Age is considered part of Power relationship: even though they are two distinct factors, putting them together allows to manipulate

Power relationship in an experiment by the participation of volunteers of different age.

d. Regionality is not considered: in study I will consider standard languages, without taking dialects into account.
e. Setting is not considered: as explained in the introduction, this factor involves the use of devices such as phone. As the experiment will be face to face and no other devices will be used, this factor can be excluded.
f. Physical distance is close enough to allow interaction: closely related to eye contact, this factor is an important one, since it limits the range of possible gestures to the ones that do not require physical contact. Supposing that in the experiment the two parties will find themselves face to face without obstacles in the middle, this factor has no longer reason to be included.
g. Gender is intended as a same-sex dyad: in sociology studies, interactions can be divided between same-sex or opposite-sex dyads. As the gender or the robot ARMAR has not been defined, the particular mechanisms of intimacy that might get triggered during opposite-sex encounters do not match the scope of this experiment. Therefore, only data regarding same-sex dyads has been considered.
h. Affect is considered together with Social Distance: this is the standard interpretation, as in Brown and Levinson's "Politeness"[92]. Using Affect as a separate variable would not match the scope of this study, which is not focused on emotions.
i. Time since last interaction is partially included in Social Distance: meeting for the first time, rather than meeting after long time, or meeting after five minutes, certainly makes a difference in the way of greeting. However, if we simplify the measurement of time, this factor becomes partially equivalent to Social Distance: "unknown person" in Social Distance equals "meeting for the first time", while "close relationship" or "acquaintance" would correspond to

"meeting after (undefined) time". Is it then convenient to act a simplification and merge this two variables into one.

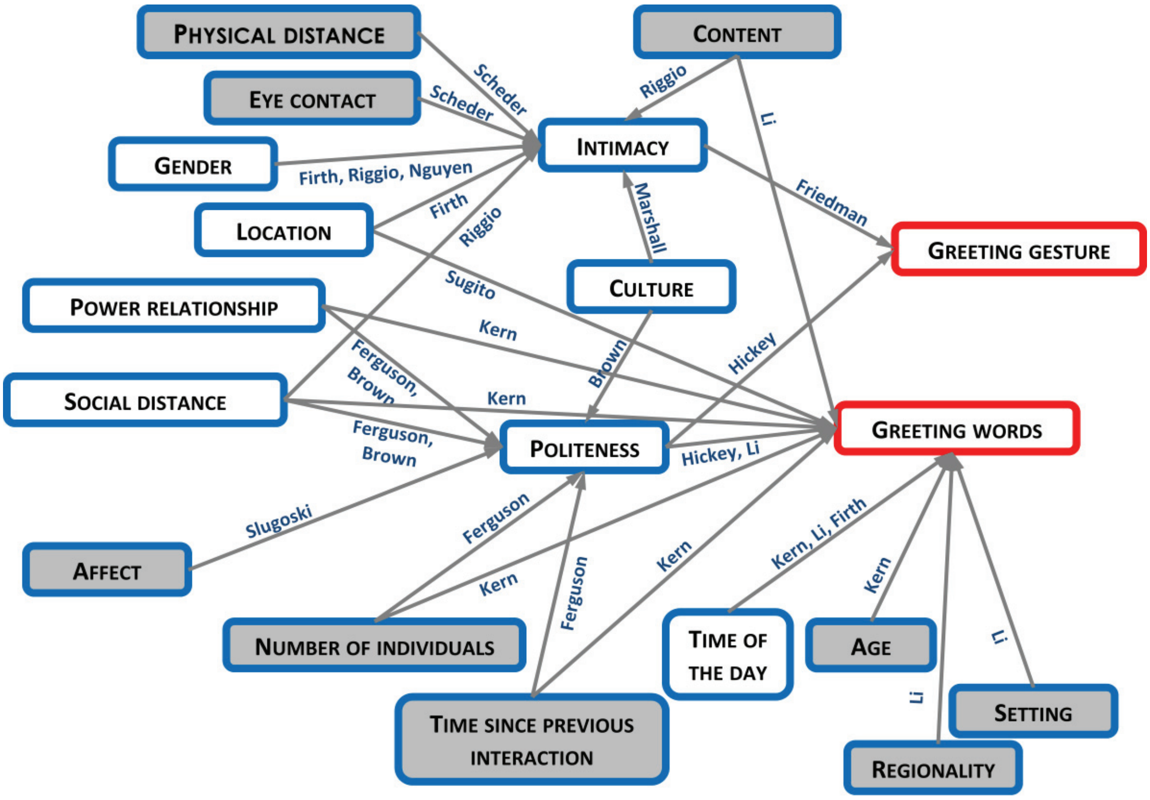


Figure 6.1 Overview of factors that influence greeting choice.
The names on the arrows indicate the authors of relevant publications.

In Figure 6.2, the graph has been simplified. It appears that Intimacy and Politeness, which are two key concepts in sociology, are intermediate passages from the upstream factors and the downstream result. For this reason, both passages can be eliminated as considered implicit in all these correlations.

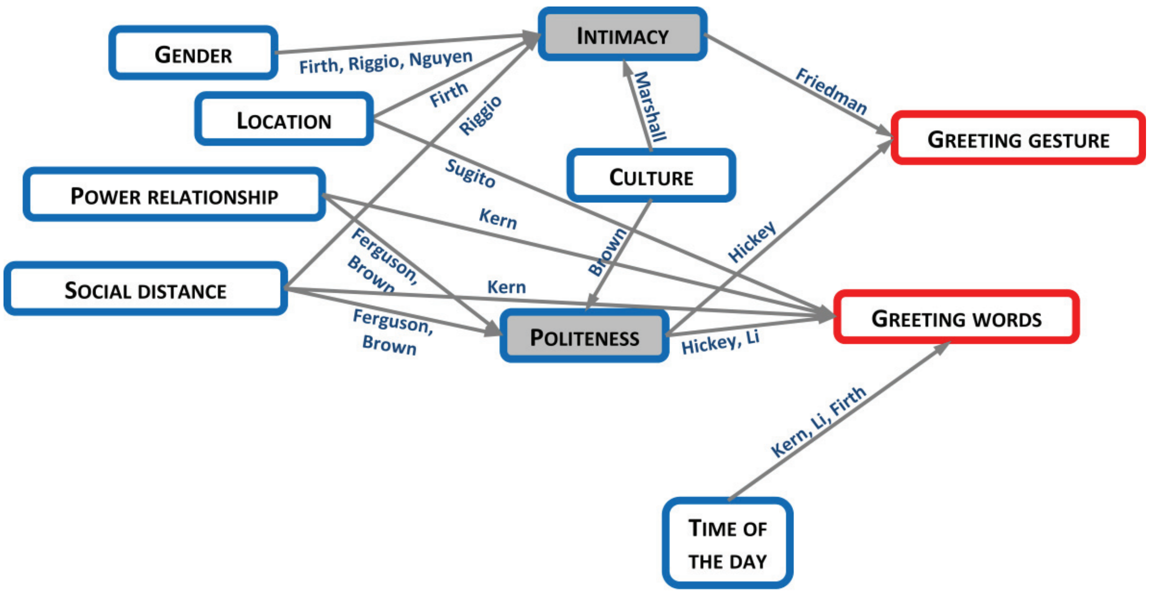


Figure 6.2 Overview of factors that influence greeting choice after simplification. The names on the arrows indicate the authors of relevant publications.

6.3 Greeting selection model

6.3.1 Overview

After the simplifications described in the previous paragraph, the resulting factors could be summarised in Figure 6.3. Among them, Culture can be considered a discriminant for switching among different mappings between the other factors and the outputs. All the other factors are then considered features of a mapping problem. They are categorical data, as they can assume only 2 or 3 values.

The outputs can also assume only a limited set of categorical values, the classes of a mapping problem. Greeting gestures list has been defined from the relevant sources mentioned in the introduction [83] [84] [85]. Originally, the set contained six gesture types, including kiss, which was dropped because not possible to implement in the robot ARMAR-IIIb, which does not have a mouth. Waving and raising a hand were also considered as broadly the same type of gesture. Greeting words list has been defined selecting the most common greeting words and getting information from relevant studies [85] [165]. This will be discussed more in depth in Paragraph 6.3.2 .

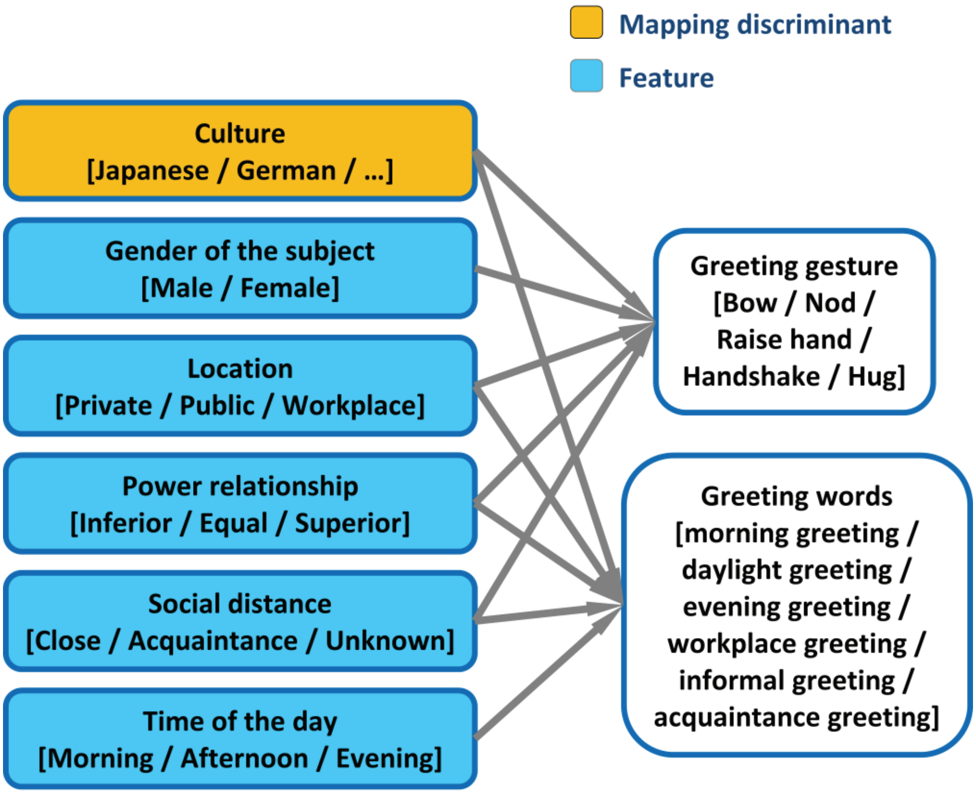


Figure 6.3 Features, mapping discriminant, outputs and their possible states.

Figure 6.4 contains the overview of the greeting model. It takes context data as input and produces the appropriate robot posture (the configuration for the chosen gesture) and speech for that input. The context is the set of features shown in Figure 6.3. Inside the mapping there is an algorithm that will be described in Paragraph 6.5 . Two different mappings are made, one for gestures and one for words. Both mapping give as output the most appropriate selection.

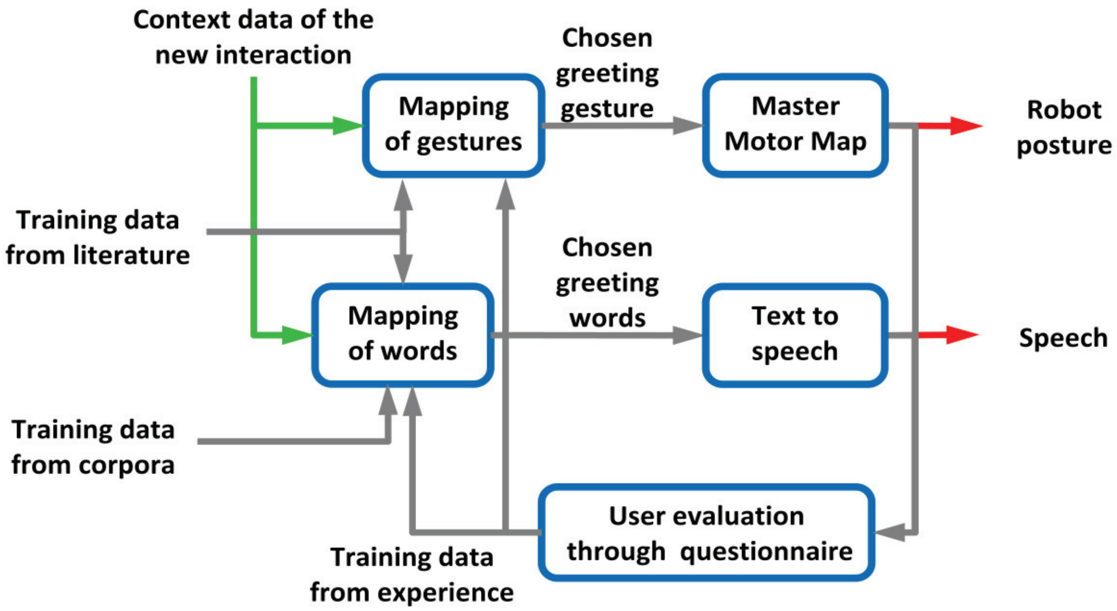


Figure 6.4 Overview of the greeting selection model.
Green arrows represent inputs; red arrows represent outputs.

In the right-hand side of the graph of Figure 6.4, these two selections are implemented. Words are turned into speech through a freeware Text-to-speech software, and the speech file is then played from the speakers of the robot. The chosen gesture is turned into robot configuration through the Master Motor Map [166], which will be described more in detail in Paragraph 6.4 .

The two outputs get evaluated by the participants of the experiment through written questionnaires. These training data that we can get from experience is given as feedback to the two mappings, which are originally trained with either data extracted from sociology studies, or in case of words, extracted from text corpora.

This model is generic: it is potentially implementable on any robot. The only robot-specific part in the present experiment is the use of Master Motor Map, which is a component that could be skipped if robot gestures are programmed manually.

6.3.2 Training data

As said in the previous paragraph, mappings can be trained with data taken from literature of sociology studies. Given the survey I made on the state of the art, it is possible to collect qualitative remarks, as well as surveys of cross-cultural studies. This data should be classified through some machine learning method or formula; nevertheless, the training data features some properties that limit the possible choice of classifying methods. Such properties are the following and lead to some considerations:

a. Incompleteness: because there are studies focused on different aspects of greeting interaction, the table we can make with is incomplete (see the "?" in the sample Table 6.1). For example, a study on gender that does not focus on other context factors provides incomplete data regarding Power relationship. Missing data make impossible to use algorithms such Principal Component Analysis or Neural Networks.
b. Cross cultural studies: some have been done in Japan, some in the U.S., some in Germany, some others compare two or more countries. Since such data can be strictly applied to the culture of the country it is referring to, the size of the training set is then strongly limited.
c. Incremental learning: as in my model I use training data from experience as feedback, the training process must be flexible to allow the mapping to change for any new data. This is also a way to overcome the problem of small training sets.
d. Supervised data: generally, sociology studies indicate that a certain gesture or word is used in a certain context. This means that learning is supervised, and excludes the use of unsupervised algorithms such as Self Organizing Maps, some of which were actually able to handle missing data.
e. Weighted data: as studies indicate a gesture or word used in a certain context, a percentage is also present in case of surveys. I added these percentages as decimal numbers between 0 and 1, like in the last column of Table 6.1. This means that each line of the training table has a

weight, and this fact excludes the use of algorithms such as Linear Discriminant Analysis, which require one single class label for each line of data.

f. Partially categorical variables: some features (such as Gender) are discrete; some other (such as Power relationship) can be discretised; while the Location (Private / Public / Workplace) for instance, is categorical. Private, Public and Workspace are three enumerates that cannot be ordered. Assigning values to them, like 0, 1, and 2, and then applying a mapping method that assumes that 0 < 1 < 2 can falsify the results.

Table 6.1 Example of training data extracted from literature

Culture	Gender	Location	Social distance	Power relationship	Gesture class	Weight
Japanese	?	1	0	?	2	0.59
Japanese	?	1	0	?	1	0.41
American	?	0	0	?	2	0.32
American	?	0	0	?	6	0.42
German	?	0	0	?	5	0.2
German	0	?	0	?	6	0.11

Considering all these limitations, I decided to use conditional probabilities: in particular Naive Bayes formula, to map the data. The Naive Bayes classifier applies Bayes' theorem with the assumption that the presence or absence of each feature is unrelated to other features. This is appropriate to the features of the present problem. Moreover, Naive Bayes only requires a small amount of training data to estimate the parameters necessary for classification. The generic formula of posterior probability is shown in Equation 4 for the class variable C_j and the features x_k from the set X.

$$p(C_j \mid X) \propto p(C_j)\prod_k p(x_k \mid C_j) \tag{4}$$

Training data of words can be obtained from text corpora as well as literature. In linguistics, a corpus is a large and structured set of texts. Sometimes portions of speech recordings get transcript into a corpus and then analysed. Corpora are then used to do statistical analysis and hypothesis testing, such as checking occurrences of a certain word in a certain context for a certain language. Conditional probabilities are calculated, like in [167], where Suprasegmental Hidden Markov Model is applied to detect emotions in speech.

It is easy to find online some corpora in English, such as the British National Corpus, or the Corpus of Historical American English. Using such online tools, it is possible to do an analysis of greeting words usage depending on the context. For example, counting the occurrence of a greeting word (Hello or Good Morning) together with some hint of distant relationship (Mr. …, Dr. ….) or close relationship (darling, etc.). In a similar way, analysis on the other features (Time of the day, Gender, etc.) can be carried out.

In Table 6.2 an example of the analysis of these data is shown. The occurrences of two words A and B are counted and their correlation is calculated. In this example, word A is “Hello”, and word B is variable (the cases of “darling” and “love” are shown): each match is shown in the first column. The other columns of the table contain respectively: the size of the whole corpus in number of words; the number of occurrences of word A; the number of occurrences of word B; the number of occurrence of both words together in a span of a number of words that is defined in the next column; the last column contains an index called Mutual Information (MI).

Table 6.2 Example of data extracted from corpora

Words A B	sizeCorpus	A	B	AB	span	MI
Hello darling	8076643	2287	600	27	1	7.31215
Hello love	8076643	2287	2553	17	1	4.55556

The index MI is defined as in equation 5. It measures the correlation between two words: the higher the value, the more the two words are strictly correlated. This value can be then normalised and it will act as a weight for the training data.

$$MI = \frac{\log\left(\frac{AB \cdot sizeCorpus}{A \cdot B \cdot span}\right)}{\log 2} \tag{5}$$

While English corpora are relatively easy to analyse, Japanese ones are more tricky due to the fact that it is often impossible for analysis tools to determine where a word ends and the next word begins, due to the lack of spaces between words in Japanese language. This fact makes calculation of span and therefore of the MI tricky or inaccurate. As a manual work of revision of huge amount of text would be necessary for solving this problem, leading to a drift away from the scope of this research, I decided to not rely on corpora for Japanese language.

Conversely, training data of Japanese greeting words was extracted from relevant Japanese sociology studies [85], [165], [168], [169], [170]. Japanese training data for gestures was extracted from [85], [171], [172], [173], while the training sets for gestures in other Western countries was made using data from [76], [83], [84].

In the present study, the location of the experiment was Germany. For this reason, the only set that was needed was the Japanese. As stated in the motivations at the beginning of this chapter, the robot should be like a foreigner: ARMAR-IIIb, trained with Japanese data, will have to interact with German people and adapt to their customs.

6.4 Implementation on ARMAR-IIIb

6.4.1 Implementation of gestures

The implementation of the set of gestures defined in Figure 6.3 on the robot was done in a way that it is not strictly hardwired to the specific hardware. Rather than defining manually the patterns of the gestures, the Master Motor Map [166] was used as intermediate passage.

The Master Motor Map (in short MMM) is a reference 3D kinematic model developed in Karlsruhe Institute of Technology, for providing a unified representation of various human motion capture systems, action recognition systems, imitation systems, visualization modules, and so on. This representation can be subsequently converted to other representations, such as action recognisers, 3D visualisation or implementation into different robots. In the framework proposed in [166] and shown in Figure 6.5, the MMM is the interface for the transfer of motion knowledge between different embodiments. The MMM is aimed to become a common standard in the robotics community, to allow having common benchmarks and sharing different software modules.

The kinematic model of MMM is expanded with statistic/anthropomorphic data, such as: segment properties (e.g. length, mass etc) defined as a function of global parameters (e.g. body height, weight). These data have been discovered and verified by various researchers, including Winter [166]. It is made setting a maximum number of DoF that might be used by any visualization, recognition, or reproduction module.

The body model of Master Motor Map based on Winter's biomechanical model [166] can be seen in Figure 6.6 on the left. It contains some joints, such as the clavicula, which are usually not implemented in humanoid robots. A converted module is necessary to perform a transformation between this kinematic model and ARMAR-IIIb kinematic model, in Figure 6.6 on the right. The converter I used [174] is a module that was created for making imitation learning tasks easier. It is based on non-linear optimisation to maximise the similarity between the demonstrated human movement and the imitation by the robot.

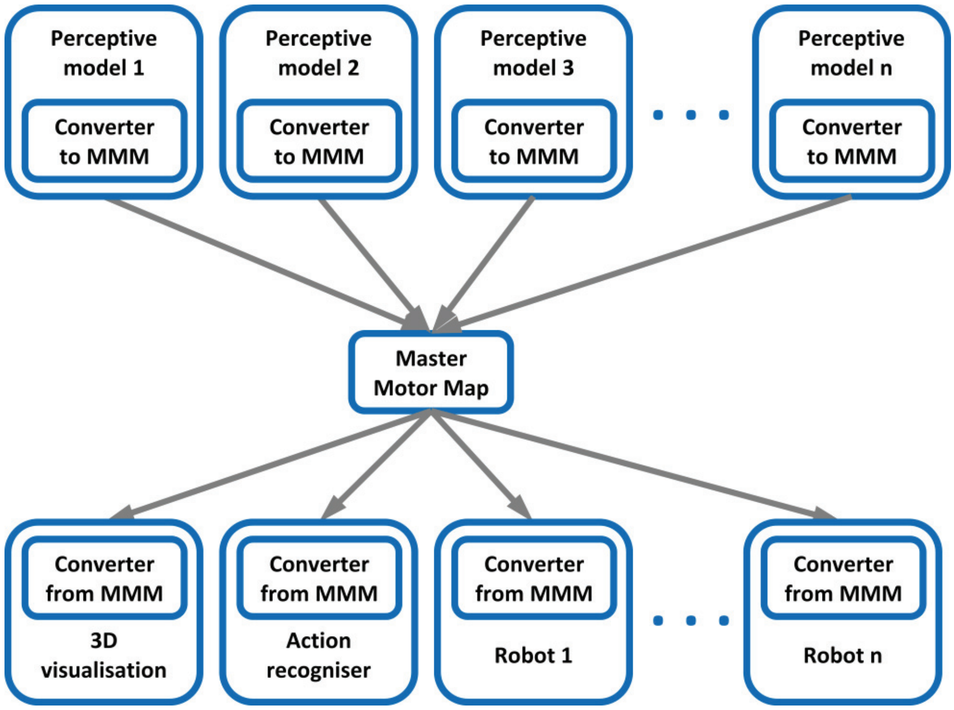


Figure 6.5 Illustration of the Master Motor Map framework.

The simplest and ideal way to reproduce a movement from given joint angles would consist in a one-to-one mapping between an observed human subject and the robot. However, due to the differences in the kinematic structures of a human and the robot (like joints and limb measurements), one-to-one mapping can hardly show acceptable results in terms of humanlike appearance of the reproduced movement. In this converter, this problem is addressed by applying a post-processing procedure in joint angle space. In two stages, the joint angles, given in the MMM format, are optimised concerning the tool centre point position and the kinematic structure of the robot through a non-linear algorithm. After a feasible solution is estimated, it serves as an initial solution for a further optimisation step, until a human-like motion on the robot is obtained.

After programming directly on the MMM model the postures (Figure 6.7), they were processed by the converter. As mentioned previously, the human model contained many joints, like pelvis, and clavicula, which are not present in the robot configuration: for instance, ARMAR cannot bend forward (for taking a bow). As there is no direct one-to-one correspondence in the joint, the conversion was not trivial.

The results we obtained with this algorithm were quite satisfying, but they needed to be retouched, due to some part of the body (e.g. the neck) not implemented in the algorithm. In Figure 6.8 the final result is shown.

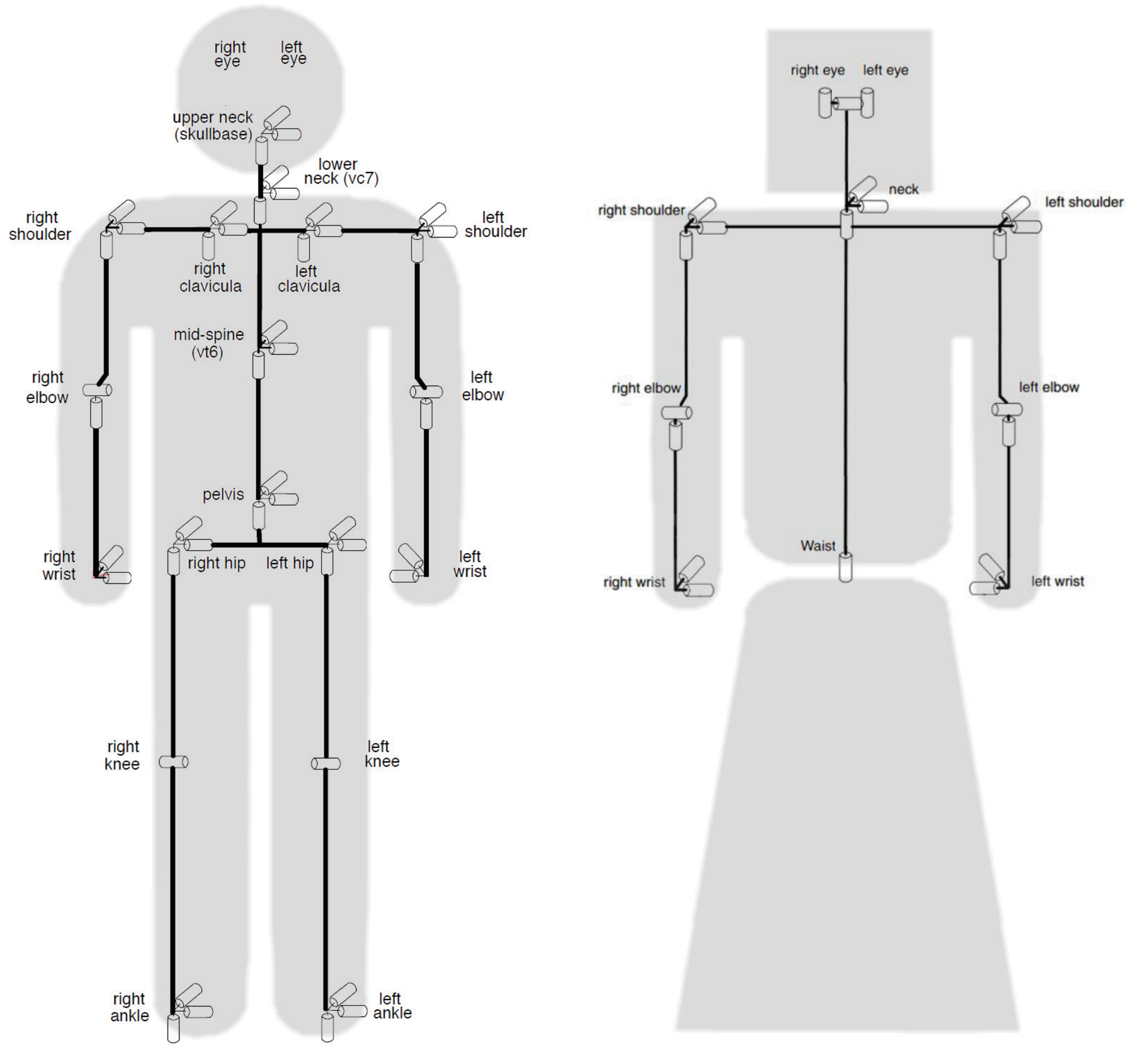


Figure 6.6 Body model of Master Motor Map and ARMAR-IIIb configuration.

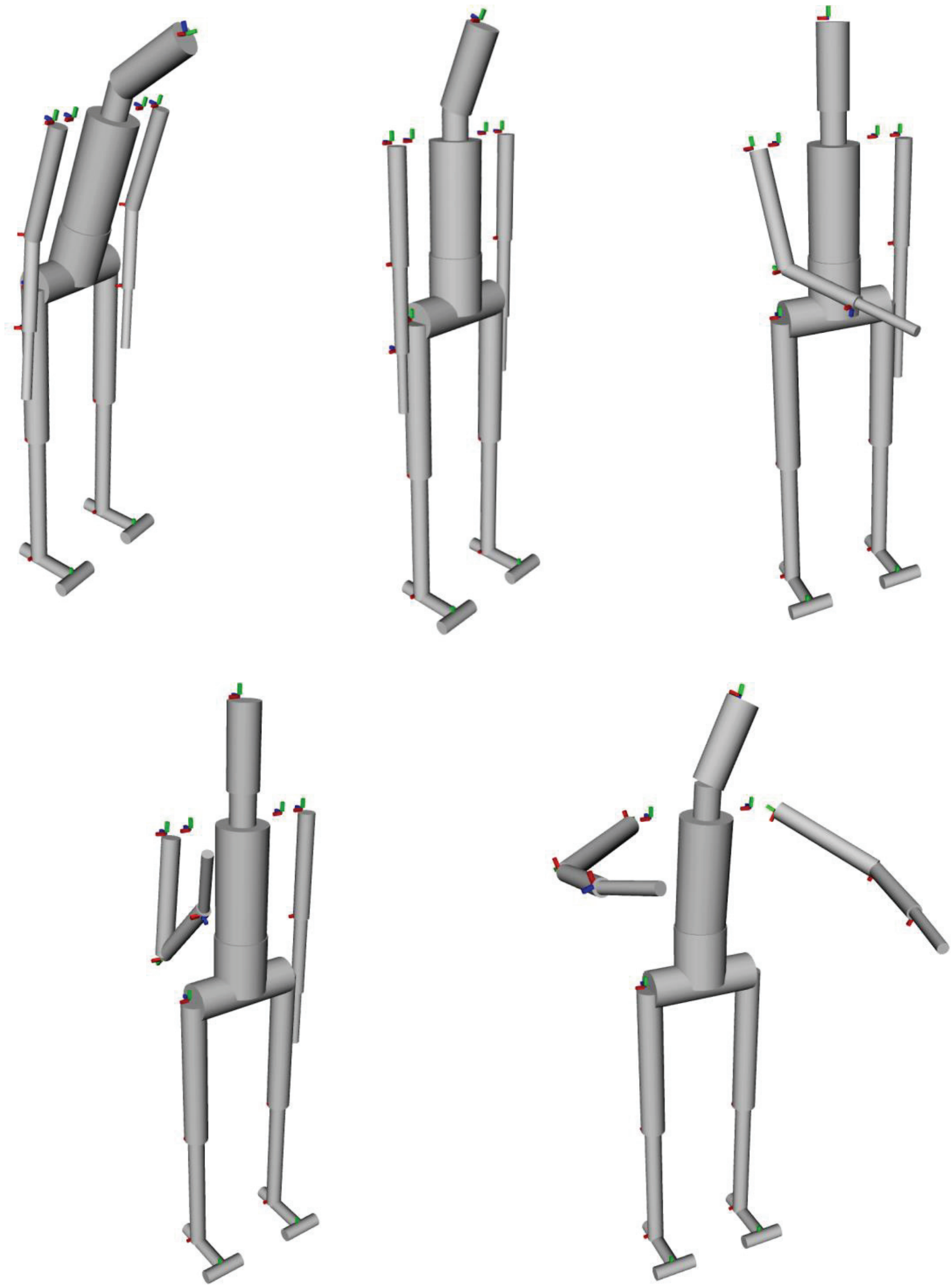


Figure 6.7 Output gestures: master motor map model.
Top row: Bow, Nod, Handshake. Bottom row: Raise hand, Hug.

Figure 6.8 Output gestures: implementation on ARMAR-IIIb.
Top row: Bow, Nod, Handshake. Bottom row: Raise hand, Hug.

The postures could be triggered from the MCA (Modular Controller Architecture, a modular software framework) interface, where the greetings model was also implemented. In Figure 6.9, the list of postures is on the left together with the option "Use Greeting Model". When that option is activated, it is possible to select the context parameters through the radio buttons on the right.

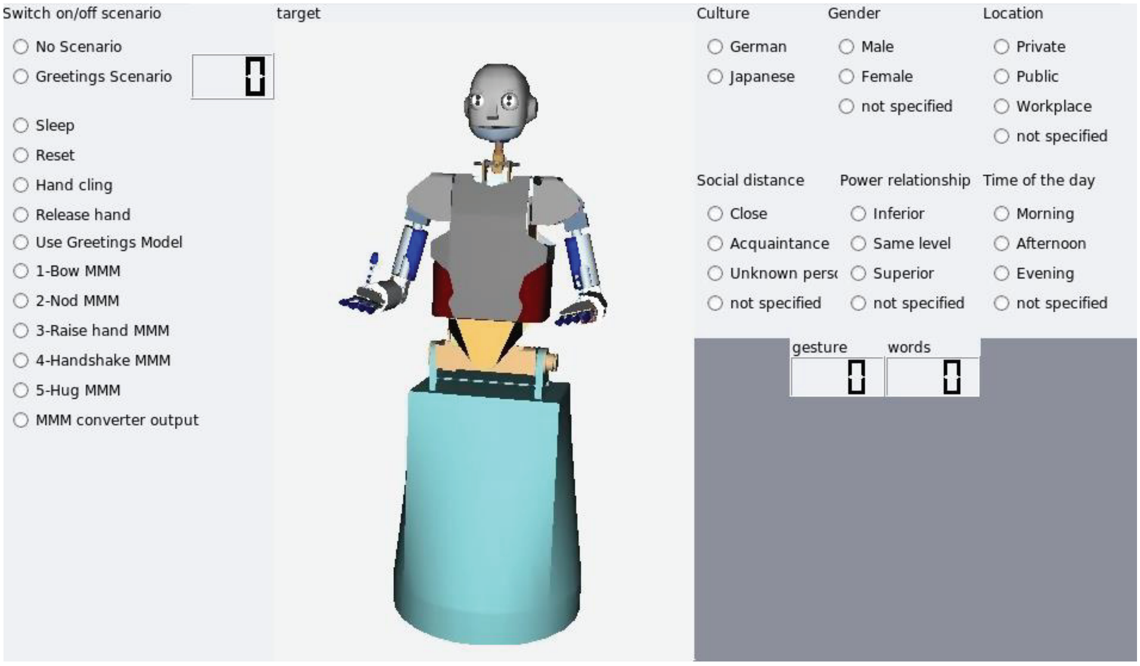


Figure 6.9 MCA Interface for the control of ARMAR-IIIb.

6.4.2 Implementation of words

As seen in Figure 6.3, the possible options of output words have been defined. This set of greetings has been translated into both German and Japanese, as in Table 6.3, regardless of the typical usage. For example, in Japan it is common to use a specific greeting in the workplace ("Otsukaresama desu"), where a standard greeting like "Konnichi wa" would be inappropriate. In German, such greeting type does not exist, but the meaning of "thank you for your effort" at work can be directly translated into German. In other words, the robot "knows a dictionary", but does not know the difference in usage of these words in different contexts.

Table 6.3 Conversion table of greeting words

Greeting type	Japanese	German
Morning greeting	Ohayō gozaimasu	Guten Morgen
Daylight greeting	Konnichi wa	Guten Tag
Evening greeting	Konban wa	Guten Abend
Informal greeting	Yō!	Hallo!
Workplace greeting	Otsukaresama desu	Vielen Dank für Ihre Mühe
Acquaintance greeting	Hajimemashite	Schön dich kennenzulernen

These words have been recorded through free text-to-speech software into wave files and could be played by the robot. ARMAR does not feature some embedded speakers in its body: for this reason, I added two small speakers behind the head and connected them to another computer.

6.5 Mapping algorithm

6.5.1 Probabilities and rewards

As mentioned in Paragraph 6.3.2 , the choice of the classification function ended up of Naive Bayes. However, mapping is not only based on that, and features an algorithm that includes rewards or penalties depending on the feedback collected from the experience (namely, participants' questionnaires). The whole algorithm is shown in Figure 6.10, and the notation contained in Table 6.4 is used.

Table 6.4 Notation used in the mapping algorithm

K	Number of elements in the dataset
$k = 1 \ldots K$	k-th element $\langle f, g_j \rangle$ in the dataset
G	Set of greetings
J	Number of possible greeting choices (5 in case of gestures; 6 for words)

$j = 1 \ldots J$	Index of greetings
g_j	j-th greeting in G
$g^{(k)}$	Greeting at the k-th element in the dataset
g^*	Greeting chosen by the classifier
g^{**}	Greeting chosen by the participant in case g^* receives a low score
I	Number of input features (4 in both cases)
$i = 1 \ldots I$	Index of features
f	Feature input vector
f_i	i-th feature of f
$f^{(k)}$	k-th feature vector in the dataset
f^*	Feature vector for the current experimental context
f^{**}	Feature vector chosen by the participant to match g^*
v	Value that can be taken by a feature f_i
$w_j^{(f)}$	Weight of the gesture j for the feature vector f
$eval$	Evaluation of the greeting g^* by the participant
r	Reward factor, depending on the evaluation of the user (for $eval \neq 3$) $r = \begin{cases} 1 & eval = 5 \\ 0.5 & eval = 4 \\ -0.5 & eval = 2 \\ -1 & eval = 1 \end{cases}$
$\delta^{(k)}$	Multiplier that varies depending on f_i. Due to the high quantity of incomplete training data, which should be assigned less importance, the multiplier in case of undefined f_i is set empirically low (0.2) $\delta^{(k)} = \begin{cases} 1 & f_i^{(k)} = v \\ 0.2 & f_i^{(k)} = -1 \\ 0 & \text{otherwise} \end{cases}$
s	Index of the participant for the current f
$l = \exp(-s/4)$	Learning factor. High (around 0.8) at the beginning and decreases following the e^{-x} curve
m	Equivalent sample size of m-estimate formula. Set empirically to 2
p	Uniform prior estimation of the probability. It is set to $1/J$

In addition to the requirements about training data discussed in 6.3.2 , the algorithm has to learn quickly. As this is a real world problem rather than an abstract one, the desired amount of iterations necessary for a complete adaptation from the initial mapping to another one should be comparable to the number of interactions human need to understand behaviour rules. The process should not require hundreds or thousands of steps.

Following Figure 6.10, let us summarise the concept of the algorithm:

1. The dataset is built from training data, with a weight $w_j^{(f)}$ for each vector corresponding to each greeting type.
2. Whenever a new feature vector f^* is given as input, it is checked whether it is already contained in the dataset or not. In the former case, the weights are directly read from the dataset; in the latter case, they get assigned the values of probabilities calculated through Naive Bayes.
3. In Naive Bayes classifier, the choice of the greeting g^* is done through a different formula from the standard one. Probabilities of each greeting g_j should be calculated taking the weights into account. However, class priors $P(g_j)$ were left out of the classifying formula. This additional assumption was made because we do not want to give more weight to more common greetings a priori, so only independent probability distributions are considered. $P(f_i|g_j)$ are balanced out through an add-ε smoothing technique, namely the use of m-estimate, which avoids the inconveniences that may happen when the number of total occurrences of a greeting under certain conditions equals 0. In generic terms, for estimating conditional probabilities in a table like the one we have, instead of using the standard formula in Equation 6, we will use the formula in Equation 7. The probability of A given B does not depend only on the joint probability of A and B divided by the probability of B, but is balanced out by p and m, which are defined in Table 6.4. We need p as a non-zero prior estimate for $P(A|B)$, which we suppose uniformly distributed over all possible values, and a number m that says how confident we are of this prior estimate p.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad (6)$$

$$P(A \mid B) = \frac{P(A \cap B) + m \cdot p}{P(B) + m} \quad (7)$$

Furthermore, as we are dealing with weights as well as missing data, a multiplier $\delta^{(k)}$ was introduced for the calculation of the joint probability of a certain feature with a certain greeting type ($\delta^{(k)}$ is defined in Table 6.4; the complete formula is inside Figure 6.10).

4. Once the greeting is chosen, the resulting probabilities get normalised. Then the stopping condition (explained in the next paragraph) is calculated. If all conditions are satisfied, no learning algorithm is performed, as the mapping has already stabilised.
5. If learning has to be performed, the first step consists in getting the evaluation from the participant. In a scale from 1 to 5, if it is greater than 3, the weight of that greeting for the present context is multiplied by a positive reward. If lesser than 3, is it multiplied by a negative reward; if it is exactly 3, nothing is done.
6. If the evaluation is lesser or equal than 3, the participant is also asked to indicate which greeting type instead would have been appropriate in this context. The weight of that greeting g^{**} is boosted by $l/2$, where the learning factor l decreases for each cycle of the algorithm. All vectors f start with a counter s set to 0, and every time one vector is processed, its counter increases as it learning factor $l = \exp(-s/4)$ decreases, leading eventually to the stabilisation of the weights. The learning factor is also used to dampen the magnitude of the reward, in the previous operation.
7. The participant is finally asked to indicate, for the chosen greeting type, which context would sound appropriate. This information becomes a new vector f^{**}, which weights are updated with a boost for the current greeting. This vector gets added to the dataset, or updated if already existing. There is another option: the participant is free to state that the present greeting should never been used. This means that there is no an f appropriate for using the present greeting. As a result, all the weights in all existing entries

in the dataset for that greeting get reduced. This action speeds up the evolution of the mapping.

8. The updated probabilities get copied into weights in the dataset and then their sum is normalised. The sum of all probabilities of greeting types for a single context combination has to be 1. At this point, the algorithm is ready for a new input (back to point 1).

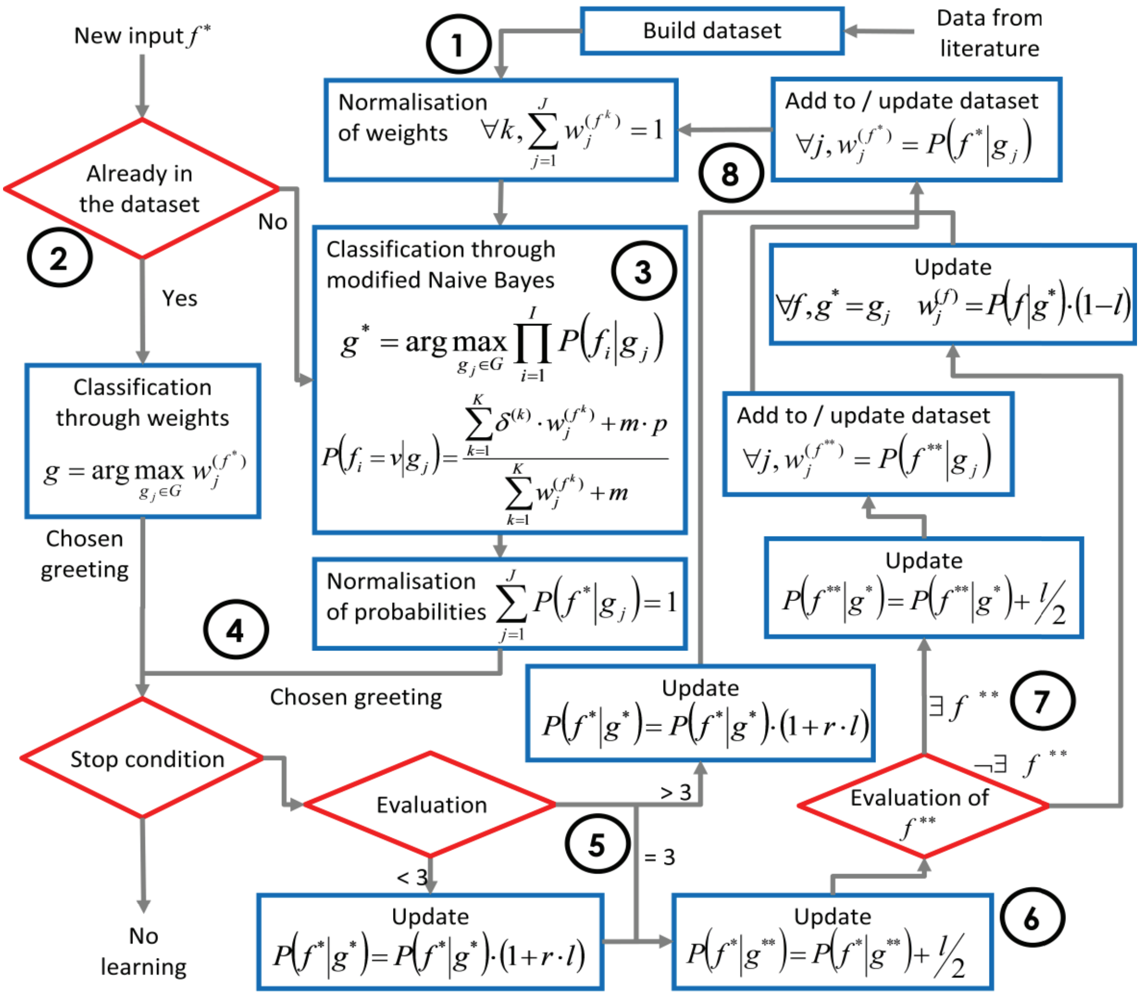


Figure 6.10 Complete overview of the mapping algorithm.

6.5.2 Stopping condition

Training stops after the state changes are stabilised. This condition is verified when the following 3 conditions (Equations 8, 9, 10; notation explained in Table 6.5) are verified at the same time. One condition ensures that all possible values of all features have been explored; the other two check whether the moving average of the latest 10 state transitions has decreased below a threshold.

$$\forall i \in I, \forall v^{(i)} \in V^{(i)}, \exists k : f_i^{(k)} = v^{(i)} \tag{8}$$

$$\frac{\sum_{t=T-9}^{T} \frac{\sigma_C^{(G)}}{\sigma_{TOT}^{(G)}}}{10} \leq \varepsilon \tag{9}$$

$$\frac{\sum_{t=T-9}^{T} \frac{\sigma_C^{(W)}}{\sigma_{TOT}^{(W)}}}{10} \leq \varepsilon \tag{10}$$

Table 6.5 Notation of the stopping conditions

$v^{(i)}$	Value of the feature i
$V^{(i)}$	Set of possible values that the feature i can take
t	Index of learning iterations
T	Current number of learning iterations
$\sigma_{TOT}^{(G)}$	Number of state changes in the greetings table
$\sigma_{TOT}^{(G)}$	Total number of states in the greetings table
$\sigma_C^{(W)}$	Number of state changes in the words table
$\sigma_{TOT}^{(W)}$	Total number of states in the words table
ε	Threshold set as 0.037, which is $2/\sigma_{TOT}^{(W)}$ (tolerance of average 2 state changes)

6.6 Experiment description

6.6.1 Participants

The experiment was performed in Germany. Participants were 18 German people of different age, gender, workplace, knowledge of the robot. The policy for recruiting consisted in covering all the possible values of each feature, as in Table 6.6, in order to ensure that the mapping could be trained with various combinations of context.

Not all combinations of feature values were possible to use in the experiment. For example, there cannot be a profile with both [Location: Workplace] and [Social distance: Unknown]. Moreover, the [Location: Private] case was left out, because it is impossible to simulate the interaction in a private context (such as one's home: the experiment took place in the laboratory). Some of the participants repeated the experiment more than once. In this way, we could collect more data just manipulating the value of one feature. A few examples:

- Social distance: a participant who meets the robot for the first time can repeat the experiment later on, and will be considered "Acquaintance" instead of "Unknown"
- Power relationship: a participant who is considered "Equal" when meeting the first time, can repeat the experiment after being explained that the robot is a puppet and trying to turn on and off the robot by themselves. In this case, the participant will be considered "Superior".
- Time of the day: the experiment is repeated at different times

The demographics of the 18 participants are as follows: M: 10; F: 8; average age: 31.33; age standard deviation: 13.16. However, the number of interactions, taking repetitions into account was 30. M: 18; F: 12; average age: 29.43; age standard deviation: 12.46. The number of participants was determined by the stopping condition of the algorithm.

Table 6.6 Participants selection policy

Features	Feature values	Profile of the participant
Gender	0: Male	Male
	1: Female	Female
Location	0: Private	–
	1: Public	People external of the laboratory where ARMAR is located
	2: Workplace	People working or studying in the laboratory
Social distance	0: Close	People who regularly work with ARMAR
	1: Acquaintance	People who have seen ARMAR before
	2: Unknown	People who have never seen ARMAR before
Power relationship	0: Inferior (ARMAR is superior to the human)	People younger than 15 OR students (profiles [Social distance: Acquaintance or Unknown]) who were told "ARMAR's cameras are taking videos that are monitored by the professor"
	1: Equal	People with no particular power relationship with ARMAR
	2: Superior (ARMAR is inferior to the human)	Professors OR elderly people OR profiles [Social distance: Acquaintance or Unknown] who were shown before the experiment that the robot is just a puppet which they can turn off pressing a button.
Time of the day	0: Morning	Any profile (the experiment is run in the morning)
	1: Afternoon	Any profile (the experiment is run in the afternoon)
	2: Evening	Any profile (the experiment is run in the evening)

6.6.2 Experimental setup

The objective of the experiment was to adapt ARMAR-IIIb greeting behaviour from Japanese to German culture. Therefore, the algorithm working for ARMAR was trained with only Japanese data taken from sociology studies (no corpora were used in the present experiment). After interacting with German people, it was expected to learn the rules of greeting interaction in Germany.

The experiment protocol is as follows:

Step 1. ARMAR-IIIb is trained with Japanese data.

Step 2. Contextual data about the encounter is given as input to the algorithm and the robot is prepared. In the meantime, the participant is instructed about what to do: enter the room, turn left and greet the robot naturally considering the current context (e.g. it's morning, in a public space, meeting for the first time, etc.).

Step 3. The participant enters the room shown in Figure 6.11. A curtain covers the location of the robot, therefore the participant will find him/herself face to face with the robot, about 2 meters distant. In this way, any possibility of greeting from distance or one of the two parties initiating greeting much before the other is avoided.

Step 4. The robot greeting is triggered by an operator as the human participant approaches (Figure 6.12). Both gestures and words get triggered at the same time. The possible choices are: [Bow / Nod / Raise hand / Handshake / Hug] and [Guten Morgen / Guten Tag / Guten Abend / Hallo! / Vielen Dank für Ihre Mühe / Schön dich kennenzulernen]. Their meaning is described respectively in Figure 6.8 and Table 6.3.

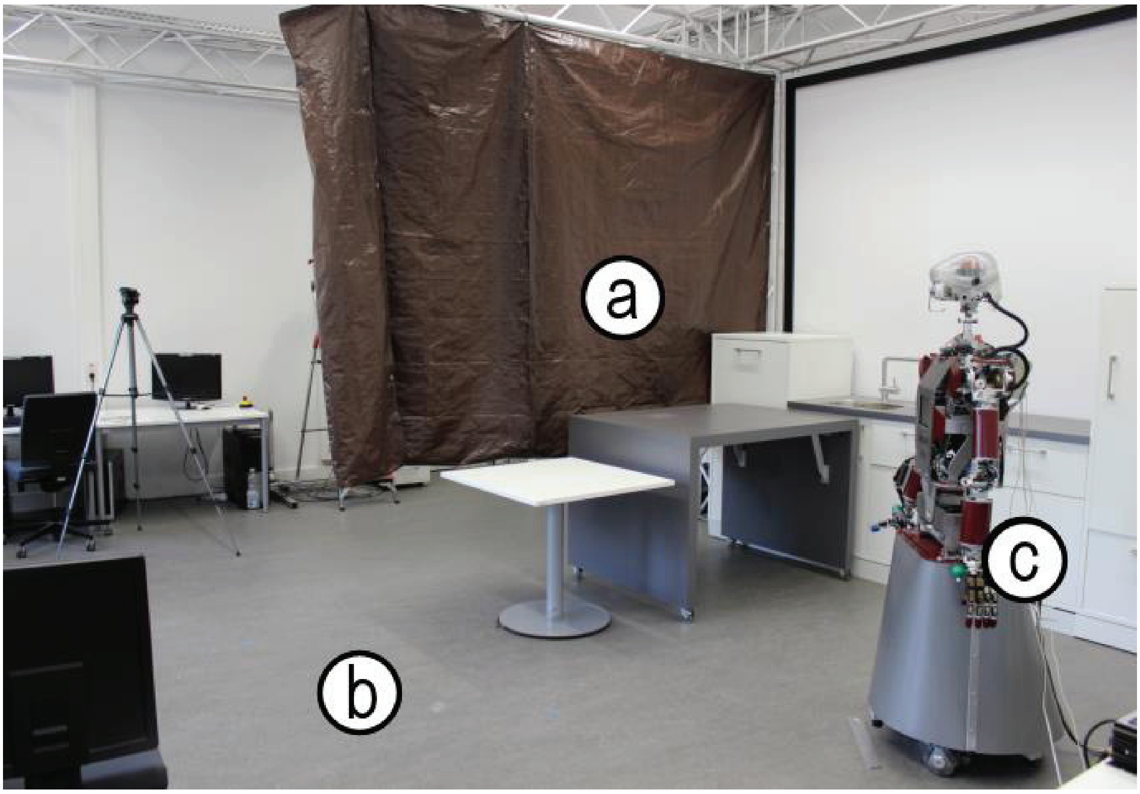


Figure 6.11 Setup of the room of the interaction experiment.
The curtain that covers the entrance (a) can be seen together with the point (b) in which the participant, after turning left, finds him/herself face to face with the robot (c).

Step 5. After the two parties have greeted each other, the robot is turned off, and the participant evaluates the robot's behaviour through a questionnaire (shown in detail in Appendix B), in order to determine:

a. whether the gesture was appropriate (in a scale from 1 to 5);

b. whether the words were appropriate (in a scale from 1 to 5);

c. in case of rating of 3 or below, which greeting gesture would have been appropriate;

d. in case of rating of 3 or below, which greeting words would have been appropriate;

e. in case of rating of 3 or below, the participant is also asked to indicate for all the input factors in which case he/she could have considered the greeting gesture or words appropriate in his/her culture, or either to specify that such gesture is never used under any condition.

Step 6. The mapping is updated using subject's feedback. The new mapping will be used in the next interaction.

Step 7. Repeat steps 2-6 for each participant.

Step 8. Training stops after the state changes are stabilised. This happens when the 3 conditions shown in Paragraph 6.5.2 are verified at the same time.

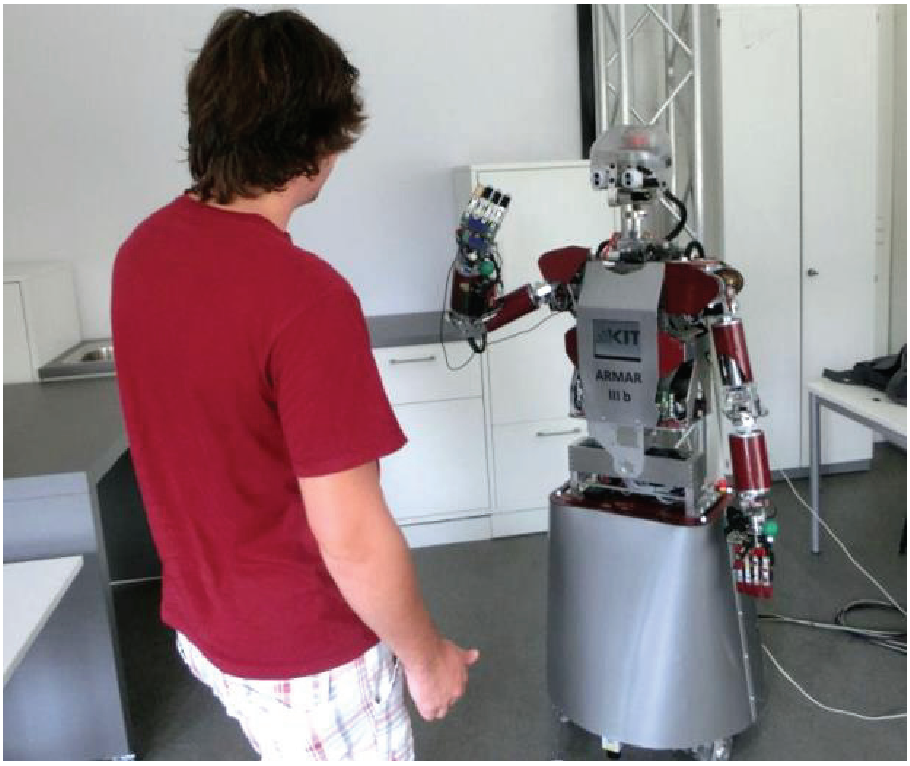


Figure 6.12 Viewpoint of the participant face to face with ARMAR.

6.6.3 Results

The experiment was carried out through 30 interactions, and all greeting gestures and word types had the chance to be selected at least once. Any behaviour mismatching with German participants' expectations did not influence their reactions, as they stuck with their own way of greeting, e.g. they would just respond raising a hand or nodding to a bow.

Handshake, shown in Figure 6.13, was common after mapping started to change. In Table 6.7 it is possible to see the evolution of the mapping of gestures. The counter T, defined already in Table 6.5 as the current number of learning iterations, corresponds to the steps 2 to 6 of the experimental protocol.

It can be noticed from the evolution of mapping that after the interactions, the amount of states in which bowing is preferred has greatly decreased, while handshake is much more spread. On the other hand, Bow has not disappeared. Hug, not present in the Japanese mapping, appears after some participant expressed their feedback indicating that hugging would be appropriate.

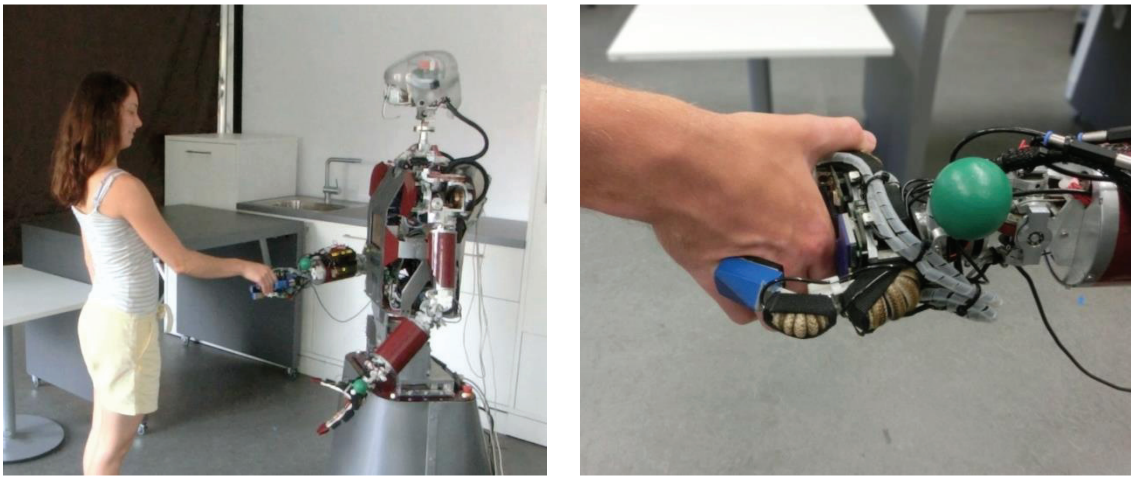


Figure 6.13 Examples of handshake with ARMAR.

Table 6.7 Evolution of mapping of gestures

T = 0		close inferior	close equal	close superior	acquain. inferior	acquain. equal	acquain. superior	unknown inferior	unknown equal	unknown superior
private	male	Nod	Raise hand	Bow	Nod	Handshake	Bow	Nod	Handshake	Bow
private	female	Nod	Raise hand	Bow	Nod	Raise hand	Bow	Nod	Bow	Bow
public	male	Nod	Raise hand	Bow	Nod	Raise hand	Bow	Nod	Raise hand	Bow
public	female	Nod	Raise hand	Bow	Nod	Raise hand	Bow	Nod	Raise hand	Bow
workplace	male	Nod	Nod	Bow	Nod	Handshake	Bow	Nod	Bow	Bow
workplace	female	Nod	Nod	Bow	Nod	Bow	Bow	Nod	Bow	Bow

T = 10		close inferior	close equal	close superior	acquain. inferior	acquain. equal	acquain. superior	unknown inferior	unknown equal	unknown superior
private	male	Nod	Nod	Bow	Nod	Handshake	Bow	Nod	Handshake	Bow
private	female	Nod	Nod	Bow	Nod	Handshake	Bow	Nod	Handshake	Bow
public	male	Nod	Raise hand	Bow	Nod	Handshake	Raise hand	Nod	Handshake	Bow
public	female	Nod	Hug	Bow	Nod	Handshake	Bow	Nod	Handshake	Bow
workplace	male	Nod	Nod	Bow	Nod	Handshake	Bow	Nod	Handshake	Bow
workplace	female	Nod	Nod	Bow	Nod	Handshake	Bow	Bow	Handshake	Bow

T = 20		close inferior	close equal	close superior	acquain. inferior	acquain. equal	acquain. superior	unknown inferior	unknown equal	unknown superior
private	male	Nod	Nod	Bow	Nod	Handshake	Handshake	Handshake	Handshake	Bow
private	female	Nod	Nod	Nod	Nod	Nod	Bow	Nod	Handshake	Bow
public	male	Raise hand	Raise hand	Raise hand	Raise hand	Handshake	Raise hand	Handshake	Handshake	Handshake
public	female	Nod	Hug	Hug	Nod	Raise hand	Handshake	Handshake	Handshake	Handshake
workplace	male	Nod	Raise hand	Handshake	Nod	Nod	Handshake	Handshake	Handshake	Handshake
workplace	female	Nod	Nod	Bow	Nod	Nod	Bow	Nod	Handshake	Bow

T = 30		close inferior	close equal	close superior	acquain. inferior	acquain. equal	acquain. superior	unknown inferior	unknown equal	unknown superior
private	male	Nod	Raise hand	Bow	Nod	Handshake	Bow	Handshake	Handshake	Bow
private	female	Nod	Nod	Bow	Nod	Nod	Bow	Nod	Handshake	Bow
public	male	Raise hand	Raise hand	Raise hand	Raise hand	Handshake	Handshake	Handshake	Handshake	Handshake
public	female	Nod	Hug	Hug	Nod	Handshake	Handshake	Handshake	Handshake	Handshake
workplace	male	Nod	Raise hand	Handshake	Nod	Nod	Handshake	Handshake	Handshake	Handshake
workplace	female	Nod	Nod	Bow	Nod	Handshake	Bow	Nod	Handshake	Bow

Bow **Nod** **Handshake** **Raise hand** **Hug**

Top row: Social distance; second row: Power relationship. Left column: Location; second column: Gender.

Table 6.8 Evolution of mapping of words

T = 0		morning inferior	morning equal	morning superior	aftern. inferior	aftern. equal	aftern. superior	evening inferior	evening equal	evening superior
private	close	Morning	Morning	Morning	Informal	Informal	Daylight	Evening	Evening	Evening
private	acquain.	Morning	Morning	Morning	Daylight	Daylight	Daylight	Evening	Evening	Evening
private	unknown	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance
public	close	Morning	Informal	Morning	Informal	Informal	Informal	Evening	Informal	Evening
public	acquain.	Morning	Morning	Morning	Daylight	Informal	Daylight	Evening	Evening	Evening
public	unknown	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Daylight	Daylight	Acquaintance	Acquaintance	Acquaintance
workplace	close	Morning	Morning	Morning	Workplace	Informal	Daylight	Workplace	Workplace	Workplace
workplace	acquain.	Morning	Morning	Morning	Workplace	Daylight	Daylight	Workplace	Workplace	Daylight
workplace	unknown	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance

T = 10		morning inferior	morning equal	morning superior	aftern. inferior	aftern. equal	aftern. superior	evening inferior	evening equal	evening superior
private	close	Morning	Morning	Morning	Informal	Informal	Daylight	Evening	Informal	Evening
private	acquain.	Morning	Morning	Morning	Daylight	Daylight	Daylight	Evening	Evening	Evening
private	unknown	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance
public	close	Morning	Informal	Morning	Informal	Informal	Informal	Evening	Informal	Evening
public	acquain.	Morning	Morning	Morning	Daylight	Informal	Daylight	Evening	Evening	Evening
public	unknown	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Daylight	Daylight	Acquaintance	Acquaintance	Acquaintance
workplace	close	Morning	Morning	Morning	Workplace	Informal	Daylight	Workplace	Informal	Workplace
workplace	acquain.	Morning	Morning	Morning	Workplace	Daylight	Daylight	Workplace	Workplace	Daylight
workplace	unknown	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance

T = 20		morning inferior	morning equal	morning superior	aftern. inferior	aftern. equal	aftern. superior	evening inferior	evening equal	evening superior
private	close	Morning	Morning	Morning	Informal	Informal	Daylight	Evening	Informal	Evening
private	acquain.	Morning	Morning	Morning	Morning	Daylight	Daylight	Evening	Evening	Evening
private	unknown	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance
public	close	Daylight	Informal	Informal	Informal	Informal	Informal	Evening	Informal	Evening
public	acquain.	Morning	Morning	Morning	Daylight	Informal	Daylight	Evening	Evening	Evening
public	unknown	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Daylight	Daylight	Acquaintance	Acquaintance	Acquaintance
workplace	close	Morning	Informal	Morning	Daylight	Informal	Evening	Evening	Informal	Evening
workplace	acquain.	Morning	Morning	Morning	Daylight	Daylight	Daylight	Evening	Evening	Evening
workplace	unknown	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance

T = 30		morning inferior	morning equal	morning superior	aftern. inferior	aftern. equal	aftern. superior	evening inferior	evening equal	evening superior
private	close	Morning	Morning	Morning	Daylight	Informal	Daylight	Evening	Informal	Evening
private	acquain.	Morning	Morning	Morning	Daylight	Daylight	Daylight	Evening	Evening	Evening
private	unknown	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance
public	close	Daylight	Informal	Informal	Informal	Informal	Informal	Evening	Informal	Evening
public	acquain.	Morning	Morning	Morning	Daylight	Informal	Daylight	Evening	Evening	Evening
public	unknown	Acquaintance	Acquaintance	Acquaintance	Daylight	Daylight	Daylight	Acquaintance	Acquaintance	Evening
workplace	close	Morning	Informal	Morning	Daylight	Informal	Daylight	Evening	Informal	Evening
workplace	acquain.	Morning	Morning	Morning	Daylight	Daylight	Daylight	Evening	Evening	Evening
workplace	unknown	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Acquaintance	Daylight	Acquaintance	Acquaintance	Acquaintance

Morning greeting | **Daylight greeting** | **Evening greeting**
Informal greeting | **Workplace greeting** | **Acquaintance greeting**

Top row: Time of the day; second row: Power relationship. Left column: Location; second column: Social distance.

Another observation is related to patterns present in the mappings: judging from the patterns in the column in Table 6.7 for T = 0, it is clearly visible that a strict categorisation is present in the Japanese mapping in regards to Social distance, whereas the same pattern is not present in the German mapping. This fact seems to go in accordance with the more hierarchical view of the society the Japanese have.

The evolution of mapping of words is shown in Table 6.8. The main change is the disappearance of the workplace greeting in German mapping, as expected.

Both resulting German and Japanese mappings may not be 100% accurate compared to reality, but they are a simplification that is consistent respectively with German participants' feedback and Japanese sociology literature.

Learning stopped after 30 iterations, when both the moving averages of state changes decreased below the threshold. The greyed out part at the left-hand side of Figure 6.14 indicates the iterations in which condition of Equation 8 was not yet true (not all feature values had been explored at least once).

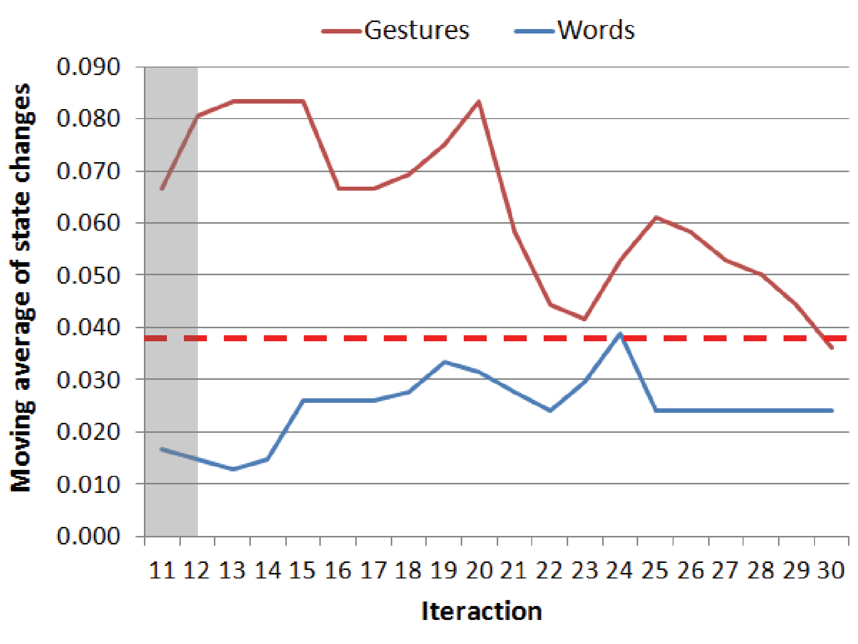


Figure 6.14 Verification of stopping criterion for each iteration.

6.7 Discussion

In this chapter, I described the development of a greeting model for humanoid robots. This work features several remarkable points worth to be mentioned.

It is the first greeting selection model ever made; it is useful for the purpose of human-robot interaction, and the comprehensive study of the state of the art of greeting choice factors is also new in sociology, together with its application in a more engineering field like robotics. In fact, existing studies focus on specific subfields of greetings: sociology studies focus on specific greetings or on the effect of specific factors; robotics studies like [175], [176] focus more on physical aspects, such as the oscillation trajectory of a handshake. My approach was more related to the mental aspect of greeting interaction, and the scope of the study was more extensive, for a less detailed greeting reproduction together with a more comprehensive view of the field.

In the described experiment, the robot can adapt to new cultures and save/load different mappings. This it is a first little step towards culture-specific robot customisation, which will be discussed more in detail in Paragraph 7.2.1 .

Another advantage of the current implementation is that gestures are not robot-specific, since the Master Motor Map framework can be used and converted to any other humanoid robot, through an either more accurate or more Naive conversion process.

Nevertheless, in the current implementation there are a few limitations, which will be described in the next paragraphs.

6.7.1 Towards a more natural interaction

The biggest limitation of this experiment is probably the need of a human operator to make the whole experiment run. The processes of inputting context variables, triggering the chosen greeting, collecting feedback data and adding it to the algorithm are all manual.

An integration with sensing abilities such as vision and speech recognition would certainly make the interaction more natural, but would be out of the scope of this specific research. Therefore, a Wizard of Oz style experiment was the most suitable solution for a research on the specific topic of greeting selection.

6.7.2 Different kinds of embodiment

Another limitation is related to a more high-level abstraction: the definition of a set of greeting gestures was done based on human related studies, taking for granted that the humanoid robot has a body that resembles humans and similar ability of motion. However, humanoid robots could be varying in shape, size and capabilities, and this could produce an effect on which greeting types are more fitting for each robot.

A possible interesting extension of this work would be making a robot find autonomously the optimal way of attracting attention and starting and interaction with a human, depending on the characteristics of its own body. Communication channels do not necessarily have to be limited to the typical human common ways of interacting, but can rely on visual or auditory aids. For instance, a device such as a mobile phone can initiating communication with a human through vibration. A robot which has blinking ability through its own eyes can use lights and colours to communicate in a dark ambient.

6.7.3 Ways of learning

The term "machine learning" was never used in this chapter, because the algorithm is made ad-hoc for the problem and does not involve complex machine learning techniques.

Even the specific implementations was chosen among other possibilities because of its feasibleness. For example, running a separate whole set of interactions for each participant would definitely help distinguishing actual learning from personal differences among participants. Having different trained mappings for different participants would also be certainly interesting for studying the distribution of the greeting type probabilities for each context.

However, due to the constraints in the participants selection policy and the manipulation of context variables, this was not possible: a male co-worker cannot do an interaction pretending to be a male acquaintance.

Further analysis on learning can be done through an additional study on a control group. A set of interactions between new German participants and the robot either using the completely trained mapping or a completely untrained one would produce interesting results, as the evaluation from the questionnaires is supposed to be significantly different. A comparative study with Japanese participant would be interesting, too.

6.8 Summary

This chapter represents the continuation of the study performed with Japanese and Egyptians. In that study, discomfort was measured in the previous experiment in case of a robot perceived as foreign. For this reason, a model for culture-dependent greeting selection was made.

A survey of the state of the art in sociology was done in the field of greetings and the resulting correlation graph was simplified to a scheme of a few context variables and one mapping discriminant. The novel model of greetings that was created features a mapping that can evolve from one culture to another and that is based on a modified version of Naive Bayes classifier. This model is generic (can be applied to any robot).

Greeting gestures and words were implemented on the humanoid robot ARMAR-IIIb and an experiment was performed with German participants. Through their feedback. ARMAR-IIIb could successfully learn a new mapping (German) of greeting selection given a defined context, starting from a Japanese mapping. This work is a step towards culture-related robots customisation.

Chapter 7

Conclusions and future work

7.1 Conclusions

7.1.1 Points of discussion

a. Realism vs usefulness

Where is the trade-off between realism of human-like appearance and the use of additional communication channels (e.g. symbols display)?

Using this additional tool, the robotic head will appear less realistic, but at the same time it will be able to communicate its emotions more clearly. This is a tool that humans do not need for communicating, because they can take advantage of the knowledge of the context, and may actually use alternative ways of expressing emotions when communication channels are limited. Humanoid robots would need additional tools to overcome their handicaps in communication.

Whether humanoid robots should be as similar as possible to humans is an open debate, and this quarrel probably resembles the one that followed the invention of photography. At that time, famous painters found themselves at a crossroads: embrace the new technology or stick to the "traditional way of painting". Eventually, photography influenced painting in stimulating new ways of capturing reality, and new painting techniques spread out, leading to a progressively stronger differentiation between the two arts. In a certain sense, new technology in robotics is stimulating us to think of design decisions in either of the two roads: more realism or more expression. In the present thesis, I have pushed in the direction of the latter one.

b. Extent of customisation

Should the customisation be by nation or by cultural area? How much language has an impact? Should the customisation include how the robot is dressed?

These open questions refer to Chapter 5. The boundary of verbal and non-verbal channels of communication could be investigated more in order to better understand the factors that link robot's behaviour and appearance with its acceptance. The first two questions will be answered after a new specific experiment, described in Paragraph 7.2.3 is carried out.

Regarding appearance, we should consider that in our experiment the two version of KOBIAN, Japanese and Egyptian were made similar on purpose: same shape, just different colour (grey instead of white). A different colour such as green would probably cause a different impression on subjects. In case of Muslim countries, this matter is even more sensitive as green is a colour commonly associated to Islam as well as Islamic political parties. On the other hand, dressing up a robot with clothes and accessories, like it was done with the robot Ibn Sina [21] or like the samurai-looking robot KIYOMORI, could be turned to an advantage for obtaining a better acceptance.

7.1.2 Limitations of this research

The main lacking part in the proposed experiments is the low degree of naturalness of the interaction. In fact, essentially all interaction are prompted by some human operator. In chapters 3 and 4 the input of emotional parameters is manual; in Chapter 5 the robots were filmed and played back with no real possibility of interaction; in Chapter 6 the interaction itself was real, but the experiment was a kind of Wizard of Oz, while input context variables and subjects' feedback were also gathered manually.

Robots need this kind of support, as it is common with impaired people. For example, a person with visual impairment may need assistance for detecting the identity of the person whom is interacting with. Robots therefore need either this kind of context information to be either inputted manually or rather be totally managed by a human experimenter.

A Wizard of Oz type of experiment, in which the robot is controlled remotely and does not have an autonomous behaviour, is also an useful tool for having a longer interaction with participants. All the studies in this thesis are limited to the beginning of an interaction: namely, the first idea that we get of robot's emotion from its expression, and the first feeling of acceptance or discomfort that we get during the greeting phase. As interaction carries on, a Wizard of Oz method becomes necessary. It can improve the affection participants feel towards the robot, as long as the trick is not exposed and expectations disappointed.

Limitations of an autonomous robots are due to hardware, as a really natural interaction would be possible only through different components elaborating auditory, visual field and so on. The potential is clear: for example KOBIAN-R, with its high number of degrees of freedom in the face, has high potential of expression, which could ideally change dynamically during an interaction by certain stimuli. It needs however more sensing abilities and integration of all these components.

7.1.3 Achievements, contributions and final remarks

Before the work of this thesis, the robot KOBIAN-R was able to perform only fixed gestures, while KOBIAN and ARMAR-III were able to perform fixed gestures. As verified, this fact caused low recognition rate of facial expressions for certain cultural groups and feeling of discomfort. Culture-adaptive system designed and developed in this thesis instead helped the robots to clearly display emotions and be in general better accepted. Some real examples of achievement of acceptance are shown in Figure 7.1 and Figure 7.2.

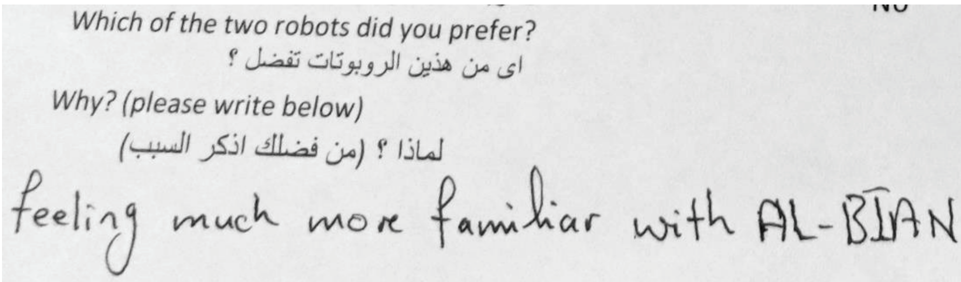

Which of the two robots did you prefer?
ای من هذين الروبوتات تفضل ؟
Why? (please write below)
لماذا ؟ (من فضلك اذكر السبب)
feeling much more familiar with AL-BIAN

Figure 7.1 Comment handwritten by an Egyptian subject stating his preference.

In the present thesis, the practical contributions to the field of research were the following four points, introduced at the beginning of the thesis, in Paragraph 1.4.2 :

a. Develop a facial expression generator and measure recognition

A generator of facial expressions was made, based on Plutchik's emotional model, on Ekman's studies for the mapping from human face muscles to KOBIAN-R's face, and achieved through classifiers and training data taken from studies by Poggi and Smith and Scott. Preliminary tests with participant from different cultures highlighted a significant recognition gap between East Asians and Westerners.

b. Develop a facial expression generator and measure recognition

The generator was used in different modes for generating "Western" or "Japanese" facial expressions, using the work of illustrators and cartoonists.

Experimental results of preferences in facial expressions and display of symbols on the robotic face are analysed, finally prompting the need for development of new hardware to make possible symbols display.

c. Study how a robot greeting is perceived depending on culture

Two cross-cultural experiments were performed and the reactions of human subjects involved in a simulated video conference with KOBIAN were observed. The subjects were either Japanese or Egyptian and the robot was greeting and speaking either like a Japanese or an Arab. The results of the investigation, whose data were gathered through Bartneck's questionnaires, suggested that Egyptians and Japanese prefer a robot adapted to their own culture, and feel symptoms of discomfort when interacting with a "foreign robot".

d. Create a model for culture dependent greeting selection

A model for culture-dependent greeting selection was made, based on sociology studies on greetings, including the works of Brown and Levinson. The model features a mapping that can evolve from one culture to another. Greeting gestures and words were implemented on ARMAR-IIIb. Through the implemented algorithm, the mapping was able to evolve from "Japanese mode" to "German mode" as ARMAR interacted with German participants.

It appears that points a and c highlighted the problem that culture-dependent discomfort and recognition gap were found. The solution came respectively in points b and d: developing flexible systems improved recognition, acceptance and interaction.

In conclusion, customising robots design and behaviour can make them perceived more familiar through different cultures.

Figure 7.2 Hug between a participant and ARMAR-IIIb.

7.2 Future work

This part will illustrate some future work. The first paragraph (7.2.1) introduces the broad direction of research; the topics presented in Paragraphs 7.2.2 , 7.2.3 and 7.2.4 instead are more practical work to be done related to different aspects of the main topic of Paragraph 7.2.1 . The last part (7.2.5) introduces an idea for a long-term project.

7.2.1 Customisation of robots

This thesis discussed about culture-related customisation. It may be argued that this is a work for its own sake, as robots are not integrated into human society yet. This might be true if we limit the analysis to nowadays robots;

however, if we think about the future and perspective development of robots, then this criticism is shortsighted.

Like technology devices are being customised for different countries [26], [27], it is reasonable to think that the more complex the product is, the more it will need customisation. Even cars are being customised for the market of different countries in small details such as the colour. In robotics, some research in this direction already exists, with the head of Flobi [69] made in modular different colours, and the prototype [177] shown in Figure 7.3.

This reasoning should be especially true for machines that resembles humans. One possible question is: should a robot have a human-like skin, or not? It is better to think in advance to such design decisions.

The present stage is thinking ahead of time about customisation for cultural groups, but the future direction consists in customisation based on: religion, language, jobs in human society, even specific customers (either thinking of the robot as the product, or the robot as a “salesman”).

The prototype in Figure 7.3 has been designed while keeping in mind that different colours will fit more some uses in the society than others, such as security rather than toy, or elderly assistance. Start thinking now about these details is important because in the future, when robots enter the mass-market, this will make a difference between a robot that gets sold and a robot that does not get sold.

Figure 7.3 Different versions of the same robotic head for different purposes.

7.2.2 Implementation of hardware for symbols display

The cultural study performed in Chapter 4 was done partially adding symbols to the photos of KOBIAN-R. Given the positive results, there is a need to determine an effective way of realising such display. The hardware solution to be implemented consists in LED panels to be placed on the top of the forehead and on the cheeks, as shown in Figure 7.4. Five devices (three on the forehead, one on each cheek) should be mounted on the new head.

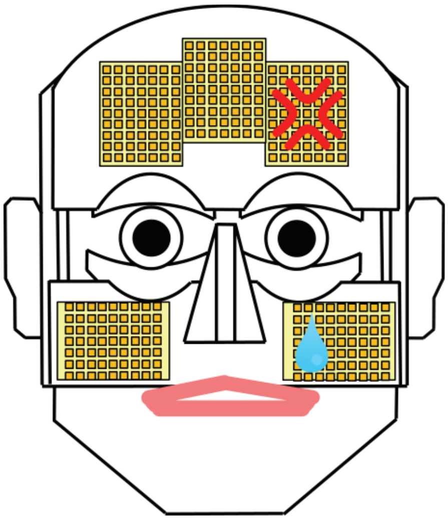
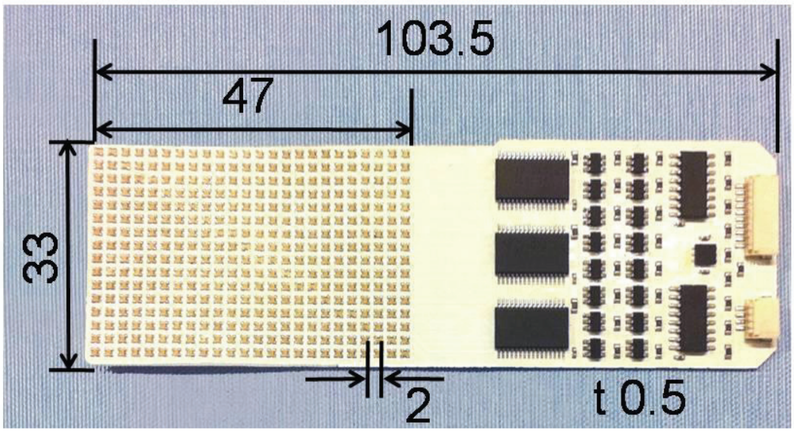


Figure 7.4 Position of LED panels to be implemented on the robotic head.

Defining dimension and position of these devices was tricky, because the required size of LED display for mark expressing area has to be not too small, while at the same time for reasons of space, the device size including wiring should be no larger than 35 x 110 mm. Due to the structure of the layers in the forehead, the device is supposed to be placed between the magnet that drives the eyebrow, the head cover, and the light diffusion sheet from one side, and the electroluminescent sheet from the other. Therefore, the thickness should be reduced as much as possible. Flexibleness is another requirement for these devices. As a result, one matrix of LED will be built on a flexible plate thinner than 1 mm. The device in Figure 7.5 has a resolution of 16 × 24, refresh rate of 10 ms and LED components of 1 x 1 mm each placed in 2 x 2 mm pitches. Brightness of each RGB colour can be controlled in 16 bit level.



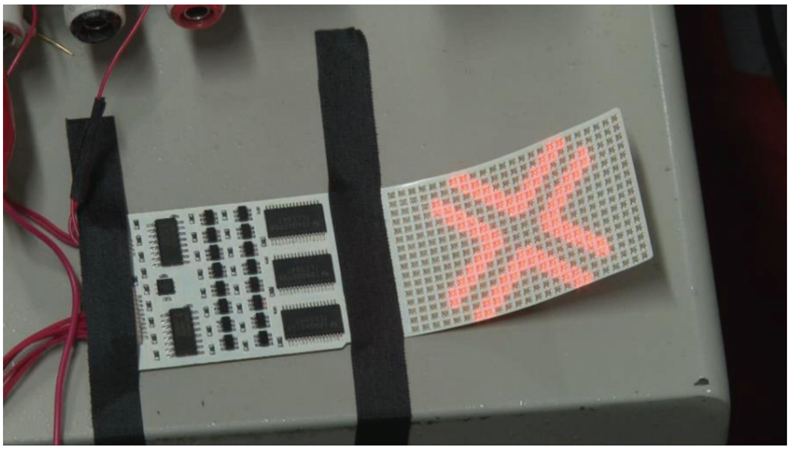


Figure 7.5 Structure of a LED panel (top picture), and example of use (bottom picture).

The solution of LED makes the realising of symbols quite flexible, as it will be possible to display of any kind of symbol or writing, not only of the small set examined in the previous study. In general, a display can be useful for many purposes in interaction: making expression recognition easier is just one of them.

In Figure 7.6 it is possible to see a preview of the final result. LED cannot however display the colour black: for that feature, some other mechanical solution will be necessary.

Further studies can be done after the hardware is completed. It would be interesting to test the symbols displayed on a neutral face, and check not only recognition but also the effect that they induce on human subjects' mental state. It is possible that for instance, the "cross-popping veins" mark makes Japanese people uncomfortable more than Westerners. Participants should be categorised beforehand depending not only on nationality but also on their familiarity with Japanese comics. Measuring discomfort could be done using the same method as the greeting experiment in Chapter 5.

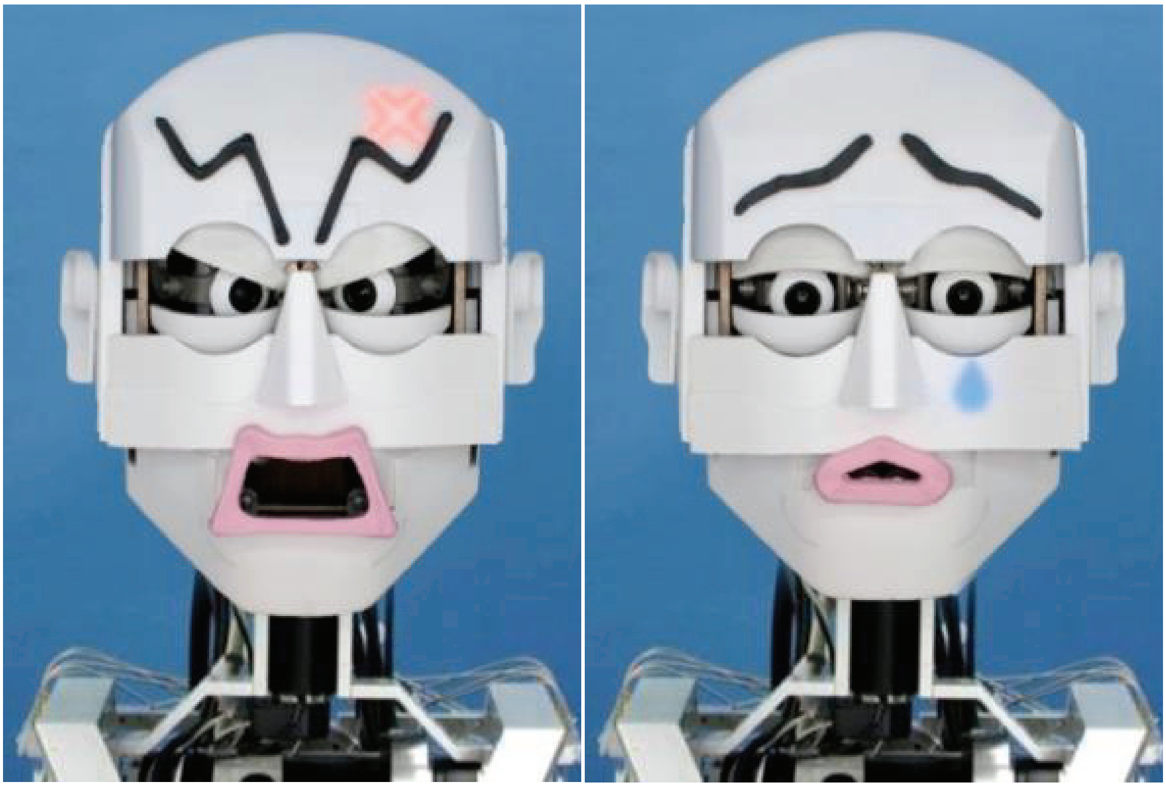


Figure 7.6 Symbols displayed using LED panels.

7.2.3 Measuring cultural distance

In Chapter 5, culture-dependent discomfort and acceptance were found in Japan and Egypt, two countries very different culturally. There is the need to repeat of the same experiment in a third Western country, far away from both.

In particular, it would be interesting, rather than repeating exactly the same experiment, to measure the effects of cultural closeness between two countries of the same area, which share similar culture but differ by language. Results we obtained in Egypt and Japan are partially related to language barrier (in Figure 5.6, the categories Understandability and Language together correspond to 42% of the reasons of preference for one robot).

Therefore, the future experiment will consist in showing two robots which will both speak a language that is not the subjects' mother tongue, respectively belonging to a culture that is considered close to the subjects' one, and to a culture that is considered distant. While words will be incomprehensible in both cases, the non-verbal channel will be, more or less familiar depending on cultural distance.

This planned experiment will take place in the Netherlands, making Dutch people interact with a German robot and a Japanese robot. If results prove that Dutch people prefer a German robot, it will support the hypothesis that customisation should be done by cultural areas; otherwise if both robots are perceived as foreigners, the customisation should be done nation by nation.

7.2.4 Greeting selection

The experiment of Chapter 6 could be extended in several ways for making the interaction more natural and overcome its limitations.

a. Manual input

One obvious limitation is related to manual input. Through the use of cameras, it would be possible to determine physical distance of the human participant and trigger the greeting automatically, rather than relying on a human operator. Computer vision would also be useful if a component for making the robot able to detect gender, age and race of the human, were developed and integrated. Context variables input would then become automatic. Feedback information could be partially made automatic through the integration of a speech recognition system and the implementation of a small dialogue.

b. Expansion of the current setup

The use of corpora of English text, which has been considered but not carried out in the experiment in Germany, could complete the extended version of this experiment, which would become a part of a bigger project, named Roboskype, described in the Paragraph 7.2.5 . Moreover, through the use of corpora, the set of possible context variables could be expanded, taking into consideration some factors that have been discarded in the simplification occurred during the definition of the greetings model (beginning of Chapter 6).

c. Strategies for initiating communication

The concept of the experiment can also be expanded, considering the limitation of the implementation of gestures itself. In fact, the set of greeting gestures was decided based on human related studies, taking for granted that the humanoid robot has a body that resembles humans and similar ability of motion. However, humanoid robots could be varying in shape, size and capabilities: they could use lights, play sounds and so on, using different channels of communication.

The optimal way of greeting may differ depending on robot features as well as context. For instance, a blinking light may be ineffective in a daylight environment whereas a sound may be unheard in a noisy environment. Gestures may also go unnoticed in a crowded environment. This reasoning leads to the conclusion that attracting humans attention is a problem that comes before greeting types. Different ways of attracting attention may also influence the perceived meaning of the message that is going to be conveyed. For example, in case of emergency, the way of initiating communication should be different from usual, in order to make the context clear to humans. In such particular case, even small details like the colour of a flashing light can make a difference.

A robot with different capabilities in terms of communication would definitely be an appropriate platform for running this king of experiments, in which depending on the context, the robot should find autonomously the optimal way of attracting attention and starting and interaction with a human, using one or more of its feature, which depend on the characteristics of its own body. Not only the optimum, but also the minimum effective set of body movements / means of communication can be investigated.

In conclusion, shape and features of a robot influence its way of communication and prompt this study, but on the other hand, the results of this study will influence in the future the design and customisation of robots for each context.

7.2.5 Roboskype

Roboskype is a project of a tele-presence system for remote communication consisting in two pairs of one human and one robot, located in different places, as in the concept in Figure 7.7. The two robots act as cultural mediators, translating not only words but also gestures and facial expressions. Preliminary studies for this concept have been done in [178].

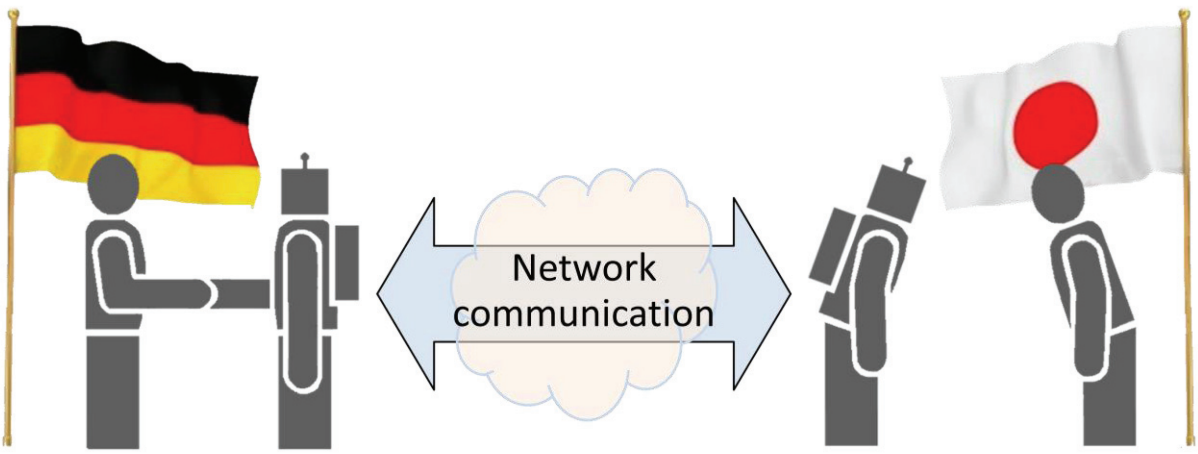


Figure 7.7 Concept of the Roboskype project.

This "cultural translation" is an ambitious goal that needs the development of several components. In this thesis, some of them that can be useful for realising this project have been developed and are shown in red border in Figure 7.8.

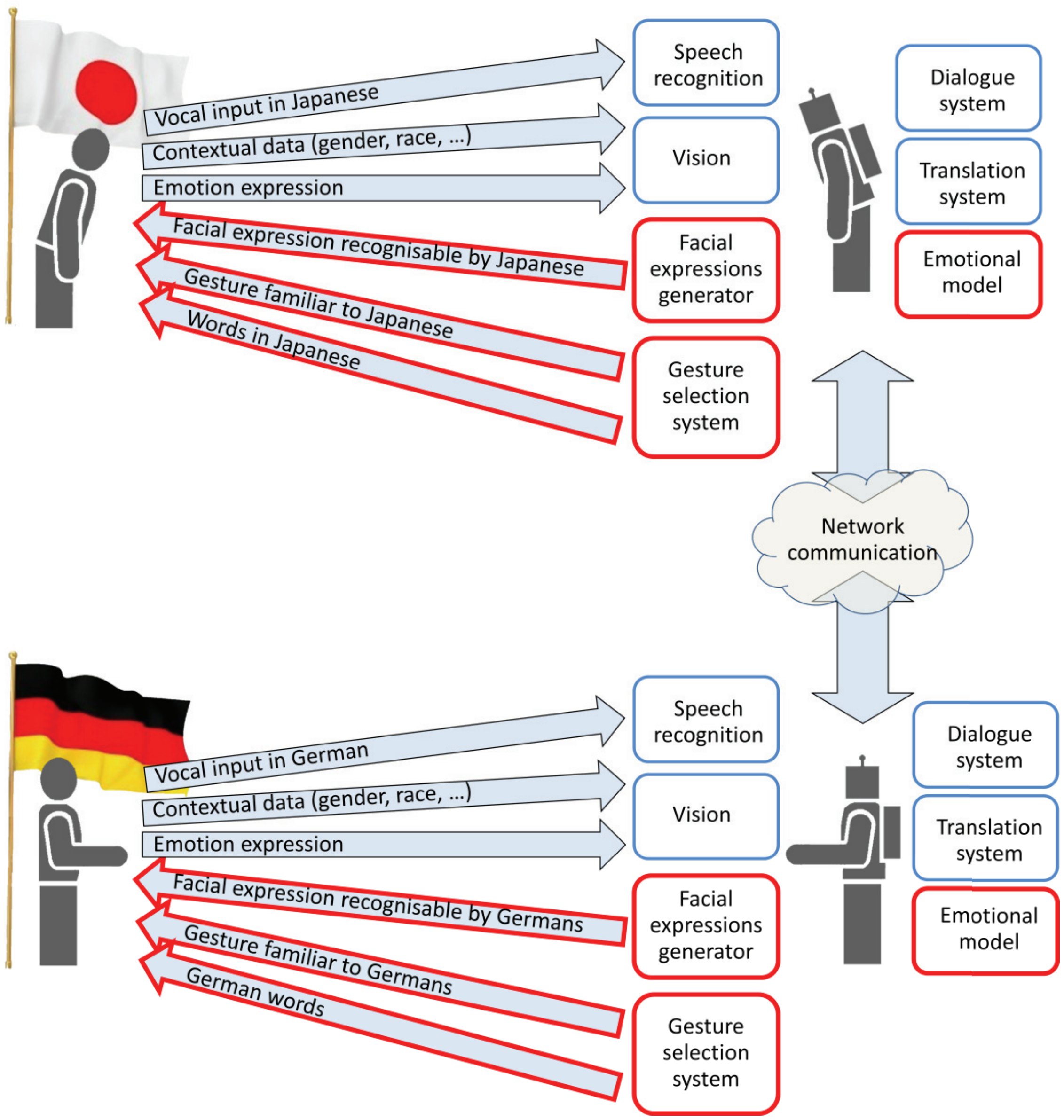


Figure 7.8 Details of the Roboskype project.
In red, components that have been developed in this thesis.

The vocal input, to be processed by a speech recognition system for the extraction of contents and voice intonation, should be integrated by other contextual data, such as distance, head orientation and hands position, measured by IMU units or Kinect, and facial expression recognised through cameras.

Through the analysis of syntax and semantics, keywords will be extracted from the message. Before the extraction, it might be necessary to modify the text, making it easier to analyse and understand. A Controlled Natural Language could be created on purpose to make a human language robot-friendly. In fact, when we talk to a foreigner, we do use a simplified version of our language. The same is also true for when we talk with a child, but the simplification is of course different. Therefore, it makes sense to think of a new simplification made for robots. Keywords are then put together with the other data in packages called "Behavioural Units", which are sent to the other side of the communication.

The receiving robot will classify the keywords and other contextual data contained in the Behavioural Unit using a database of lexicon of non-verbal communication, which is stored in a "cloud" of knowledge available to both robots. Written text will instead be converted into the remote native language by an automatic translation system and then processed by a text-to-speech.

The output contextual data (facial expression, gestures, voice intonation and so on) will be classified through different (culture-dependent) training rules at the two sides of the communication. This will result in an output non-verbal expression specific to the local culture.

Finally, the response of the robot will be specific for the machine's architecture, like in case of different body shape or number of degrees of freedom, the specific implementation of gestures and facial expressions will change.

Appendix A

Random facial expressions

In this part of the appendix the facial expressions obtained by using random vectors are shown. For each figure it is possible to see the output expression together with the input, giving some hints on the effect of the contribution of each component of the input vector. The 12 basic vectors are shown in Table A.1 for reference.

Table A.1 Components of the vectors of basic emotions and communication acts

	Mood	**Stance**	**Temperament**	**Expectation**	**Certainty**	**Power Rel.**
Neutral	0	0	0	0	0	0
Happiness	66	0	0	0	0	0
Trust	0	66	0	0	0	0
Fear	0	0	66	0	0	0
Surprise	0	0	0	66	0	0
Comprehension	0	0	0	0	66	0
Superiority	0	0	0	0	0	66
Sadness	-66	0	0	0	0	0
Disgust	0	-66	0	0	0	0
Anger	0	0	-66	0	0	0
Anticipation	0	0	0	-66	0	0
Incomprehension	0	0	0	0	-66	0
Inferiority	0	0	0	0	0	-66

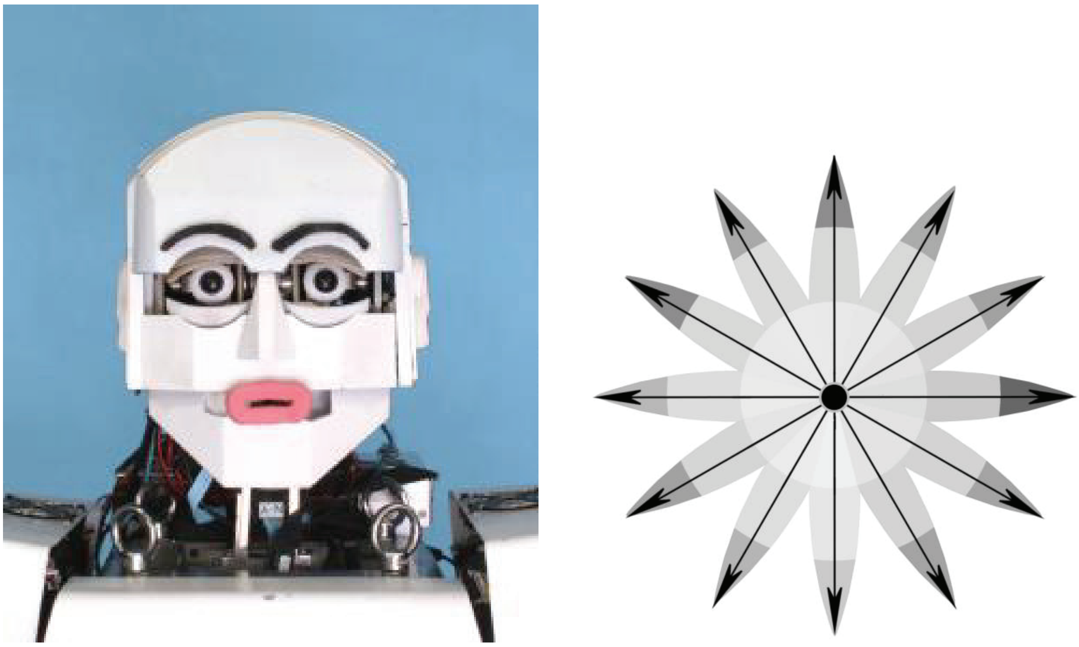


Figure A.1 Facial expression produced from neutral vector (0, 0, 0, 0, 0, 0).

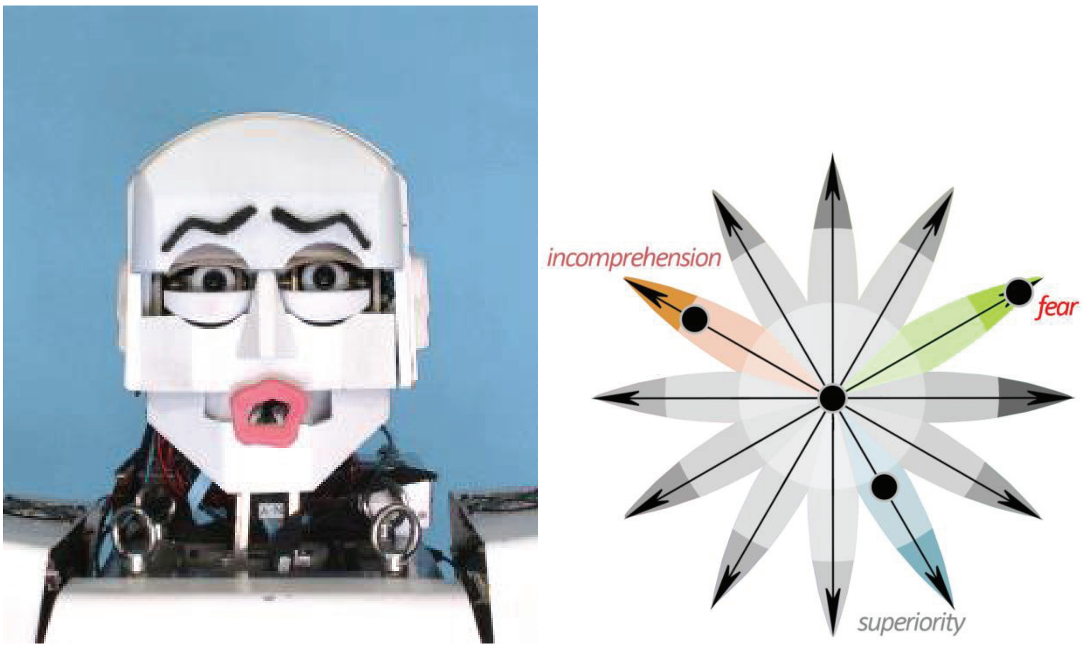


Figure A.2 Facial expression produced from vector (0, 0, 95, 0, -62, 35).

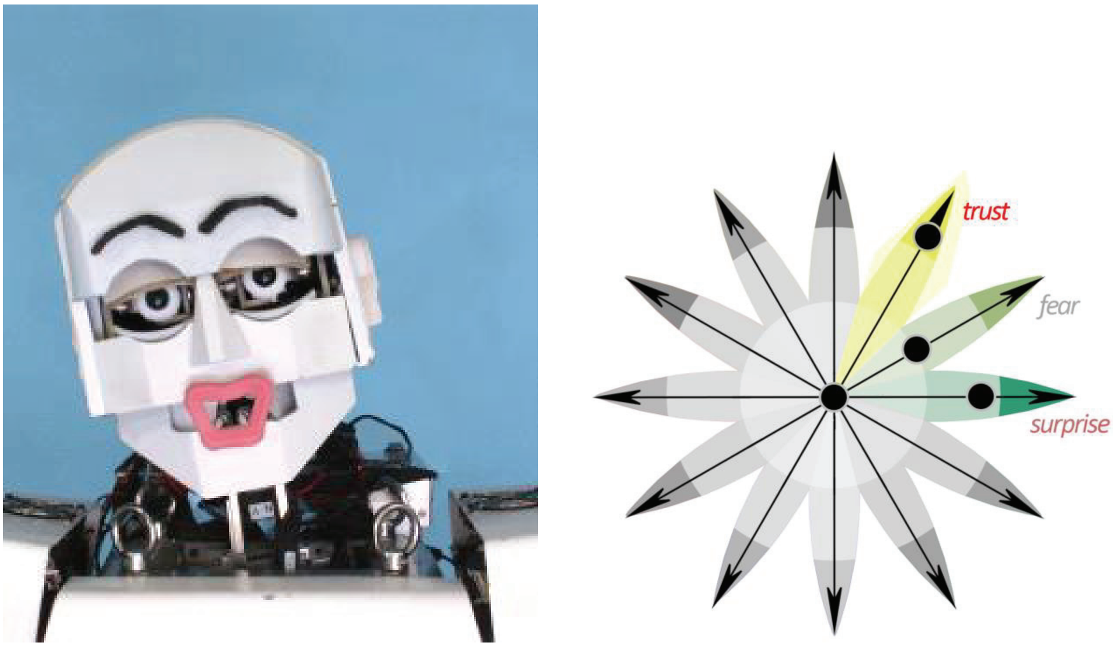


Figure A.3 Facial expression produced from vector (0, 70, 39, 56, 0, 0).

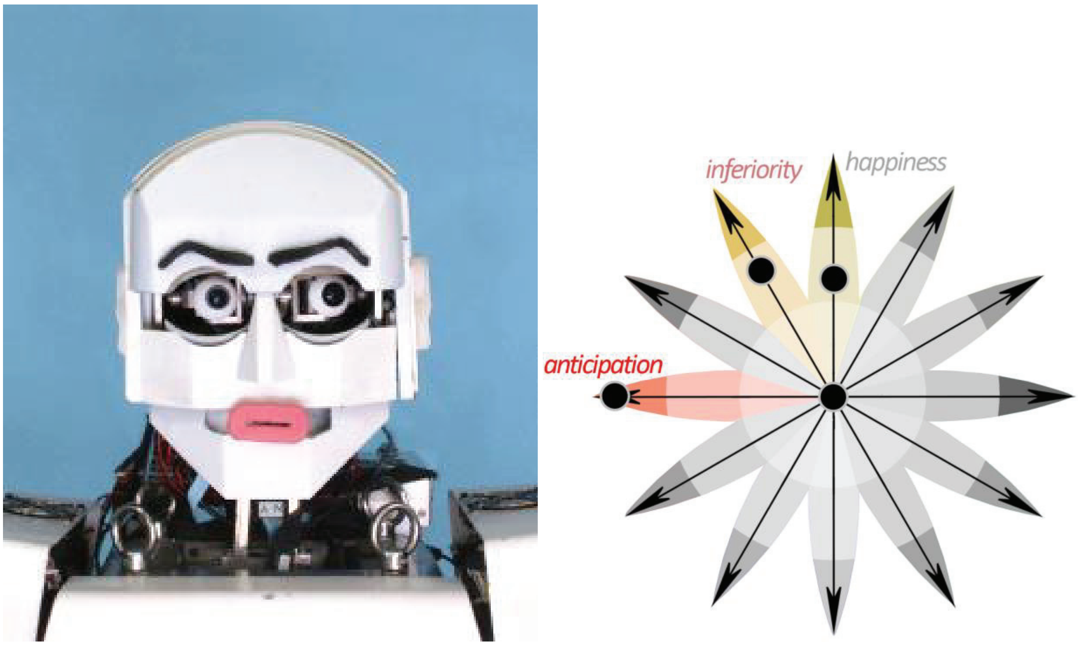


Figure A.4 Facial expression produced from vector (51, 0, 0, -89, 0, -60).

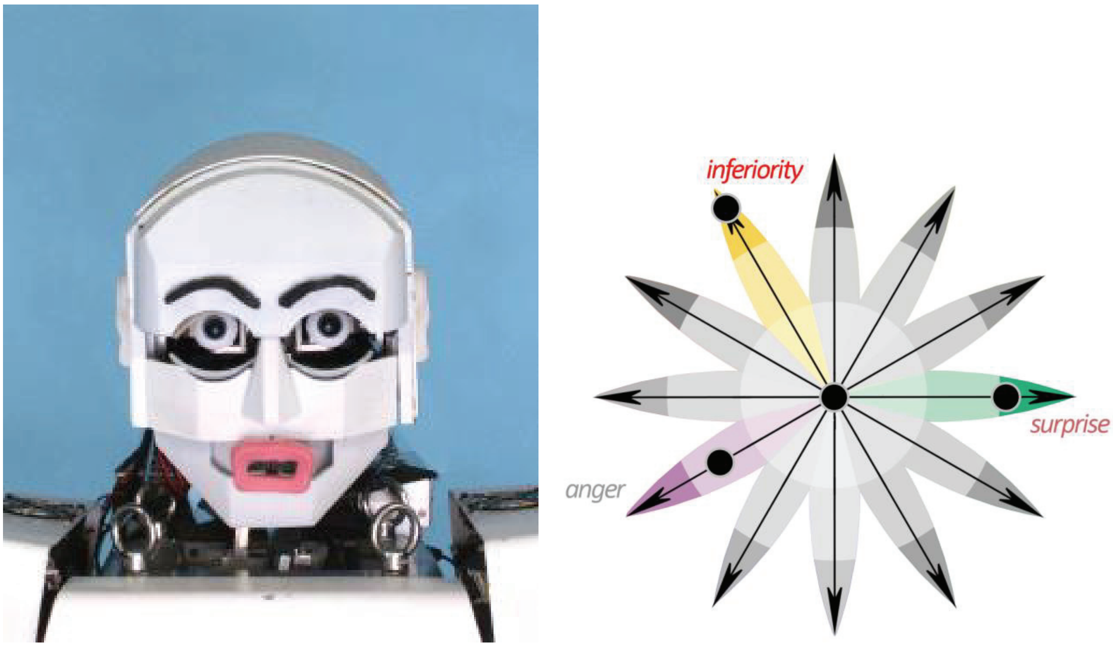


Figure A.5 Facial expression produced from vector (0, 0, -58, 68, 0, -87).

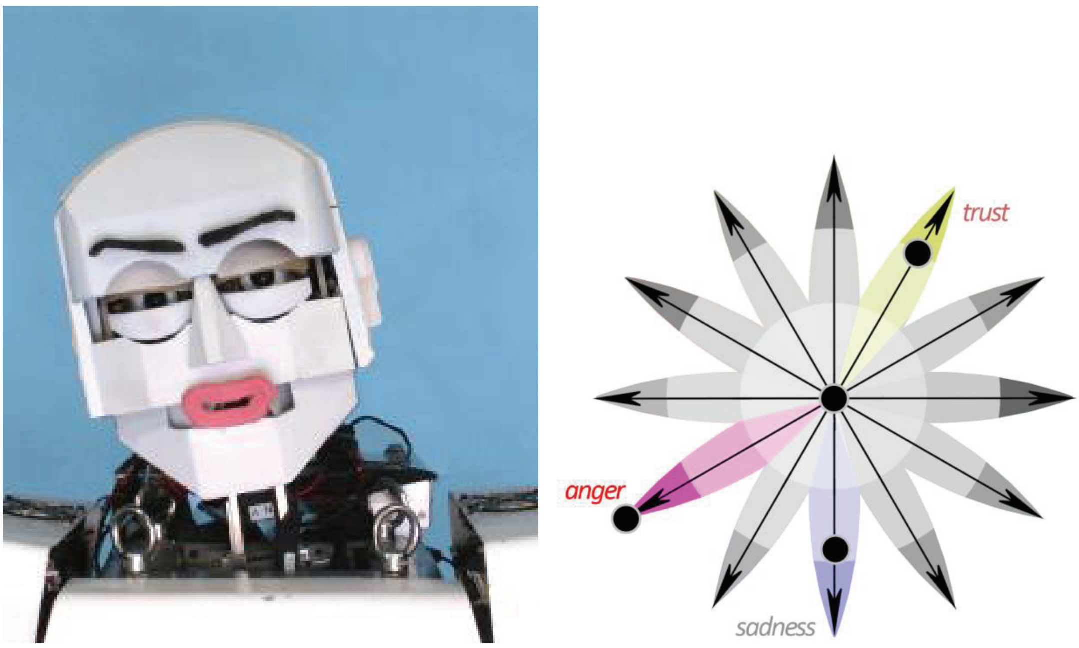


Figure A.6 Facial expression produced from vector (-60, 67, -98, 0, 0, 0).

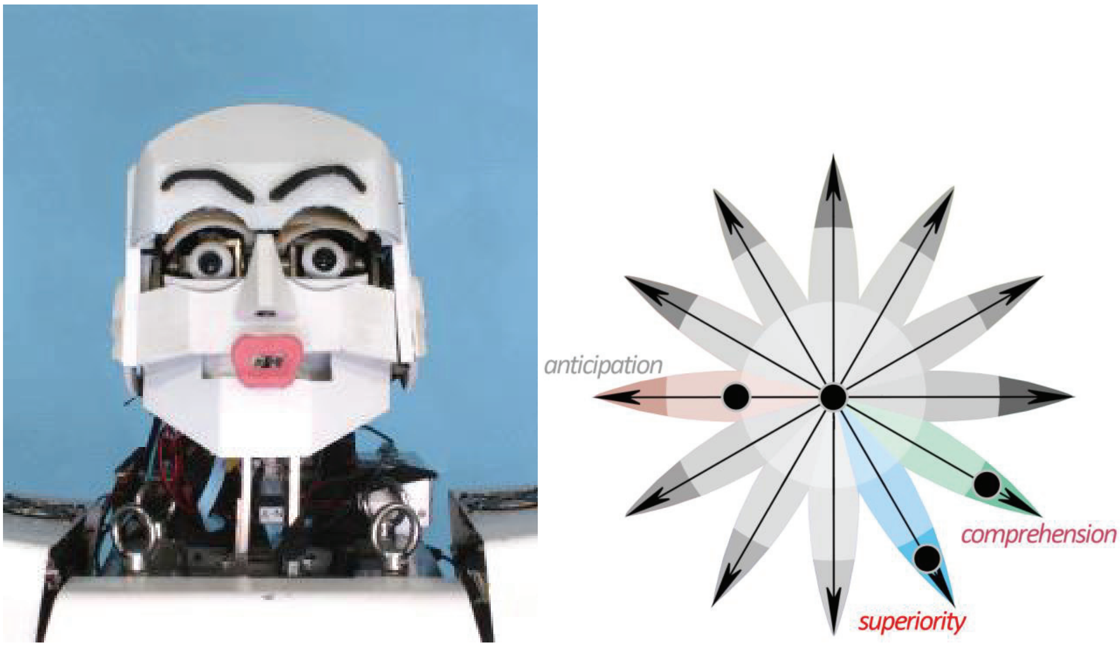


Figure A.7 Facial expression produced from vector (0, 0, 0, -46, 77, 79).

Appendix B

Questionnaires

In this part of the appendix the materials used in the experiment described in chapters 5 and 6 are shown.

B.1 Questionnaire used in Egypt

This questionnaire was filled at the beginning of the experiment (Figure B.1); after the first interaction, with KOBIAN (Figure B.2); and after the second interaction, with AL-BIAN (Figure B.3). The last part (Figure B.4) was used by the experimenter.

In case the two robots were presented in the opposite order (AL-BIAN first, then KOBIAN), the questionnaire of Figure B.3 was used before the one of Figure B.2, with the exception of the last question "Which if the two robots did you prefer?", which should always be at the end of the experiment.

Questionnaire about humanoid robots

Your sex? الجنس؟ — Male ذكر — Female أنثى

Are you familiar with humanoid robots? هل انت على دراية عن الروبوتات ؟ — Yes نعم — No لا

Your age? العمر؟ ☐

Your nationality? البلد؟ — Egypt — Japan — Other (please write)

Your education level? المؤهل العلمي

Below high school	أقل من الثانوية العامة	☐
High school	الثانوية العامة	☐
Bachelor's degree	درجة البكالوريوس	☐
Master's degree	شهادة ماجستير	☐
PhD or more	شهادة دكتوراه	☐

Please rate your thoughts about humanoid robots on those scales:

من فضلك حدد تقيمك عن الانسان الالى

Dislike	غير معجب	1	2	3	4	5	Like	معجب
Unfriendly	غير صديق	1	2	3	4	5	Friendly	صديق
Unkind	غير طيب	1	2	3	4	5	Kind	طيب
Unpleasant	غير ممتع	1	2	3	4	5	Pleasant	ممتع
Awful	غير ظريف	1	2	3	4	5	Nice	ظريف

Please rate your emotional state on these scales:

من فضلك قيم انطباعك على الروبوت من خلال المفاهيم الاتية

Anxious	قلق	1	2	3	4	5	Relaxed	مريح للأعصاب
Agitated	مرتبك	1	2	3	4	5	Calm	هادى
Quiescent	غير متفاجئ	1	2	3	4	5	Surprised	متفاجئ

Figure B.1 Part 1 of the questionnaire used in Egypt.

Questionnaire about the robot KOBIAN
استطلاع عن الروبوت KOBIAN

Please rate your impression of the robot on these scales:

من فضلك قيم انطباعك على الروبوت من خلال المفاهيم الاتية

Dislike	غير معجب	1	2	3	4	5	Like	معجب
Unfriendly	غير صديق	1	2	3	4	5	Friendly	صديق
Unkind	غير طيب	1	2	3	4	5	Kind	طيب
Unpleasant	غير ممتع	1	2	3	4	5	Pleasant	ممتع
Awful	غير ظريف	1	2	3	4	5	Nice	ظريف

Please evaluate the robot's behaviour on these scales:

من فضلك قيم اسلوب الروبوت من خلال المفاهيم الاتية

Impolite	غير مهذب	1	2	3	4	5	Polite	مهذب
Mysterious	غريب	1	2	3	4	5	Familiar	قريب للقلب
Incomprehensible	صعب الفهم	1	2	3	4	5	Comprehensible	سهل الفهم
Foreign	غريب	1	2	3	4	5	Native	محلى

Please rate your emotional state on these scales:

من فضلك قيم انطباعك على الروبوت من خلال المفاهيم الاتية

Anxious	قلق	1	2	3	4	5	Relaxed	مريح للأعصلب
Agitated	مرتبك	1	2	3	4	5	Calm	هادى
Quiescent	غير متفاجئ	1	2	3	4	5	Surprised	متفاجئ

Did you like the way the robot speaks English? هل أعجبتك طريقة تحدث الروبوت باللغة الانجليزية ؟	No	لا	1	2	3	4	5	نعم	Yes
Did you like the gesture the robot uses to greet you? هل أعجبتك طريقة تحية الروبوت لك ؟	No	لا	1	2	3	4	5	نعم	Yes
Did you like the gesture the human operator uses to greet you? هل أعجبتك طريقة تحية الرجل لك ؟	No	لا	1	2	3	4	5	نعم	Yes
Did you like the greeting speech ("Konnichi wa") of the robot? هل أعجبك الروبوت اثناء تحيتة بكلمة "كوننيتشيوا" لك ؟	No	لا	1	2	3	4	5	نعم	Yes
Did you like the greeting speech ("Konnichi wa") of the human operator? هل أعجبك الرجل اثناء تحيتة بكلمة "كوننيتشيوا" لك ؟	No	لا	1	2	3	4	5	نعم	Yes
Would you like to meet the robot again? هل تريد مقابلة الروبوت مرة أخرى؟	No	لا	1	2	3	4	5	نعم	Yes

Figure B.2 Part 2 of the questionnaire used in Egypt.

Questionnaire about the robot AL-BIAN
استطلاع عن الروبوت AL-BIAN

Please rate your impression of the robot on these scales:

من فضلك قيم انطباعك على الروبوت من خلال المفاهيم الاتية

Dislike	غير معجب	1	2	3	4	5	Like	معجب
Unfriendly	غير صديق	1	2	3	4	5	Friendly	صديق
Unkind	غير طيب	1	2	3	4	5	Kind	طيب
Unpleasant	غير ممتع	1	2	3	4	5	Pleasant	ممتع
Awful	غير ظريف	1	2	3	4	5	Nice	ظريف

Please evaluate the robot's behaviour on these scales:

من فضلك قيم اسلوب الروبوت من خلال المفاهيم الاتية

Impolite	غير مهذب	1	2	3	4	5	Polite	مهذب
Mysterious	غريب	1	2	3	4	5	Familiar	قريب للقلب
Incomprehensible	صعب الفهم	1	2	3	4	5	Comprehensible	سهل الفهم
Foreign	غريب	1	2	3	4	5	Native	محلى

Please rate your emotional state on these scales:

من فضلك قيم انطباعك على الروبوت من خلال المفاهيم الاتية

Anxious	قلق	1	2	3	4	5	Relaxed	مريح للأعصاب
Agitated	مرتبك	1	2	3	4	5	Calm	هادى
Quiescent	غير متفاجئ	1	2	3	4	5	Surprised	متفاجئ

Question										
Did you like the way the robot speaks English? هل أعجبتك طريقة تحدث الروبوت باللغة الانجليزية ؟	No	لا	1	2	3	4	5	نعم	Yes
Did you like the gesture the robot uses to greet you? هل أعجبتك طريقة تحية الروبوت لك ؟	No	لا	1	2	3	4	5	نعم	Yes
Did you like the gesture the human operator uses to greet you? هل أعجبتك طريقة تحية الرجل لك ؟	No	لا	1	2	3	4	5	نعم	Yes
Did you like the greeting speech ("Alsalamo alikum") of the robot? هل أعجبك الروبوت اثناء تحيتة بكلمة "السلام عليكم" لك ؟	No	لا	1	2	3	4	5	نعم	Yes
Did you like the greeting speech ("Alsalamo alikum") of the human operator? هل أعجبك الرجل اثناء تحيتة بكلمة "السلام عليكم" لك ؟	No	لا	1	2	3	4	5	نعم	Yes
Would you like to meet the robot again? هل تريد مقابلة الروبوت مرة أخرى؟	No	لا	1	2	3	4	5	نعم	Yes

Question							
Which of the two robots did you prefer? اى من هذين الروبوتات تفضل ؟	KOBIAN	1	2	3	4	5	AL-BIAN

Why? (please write below)

لماذا ؟ (من فضلك اذكر السبب)

Figure B.3 Part 3 of the questionnaire used in Egypt.

Questionnaire for internal use

	Operator (JAP)	Operator (ARA)	KOBIAN	AL-BIAN
Answer to initial greeting (spoken)	☐	☐	☐	☐
response time?	....	....	....	....
matching? (Yes / No)	....	....	....	....
Answer to initial greeting (gesture)	☐	☐	☐	☐
response time?	....	....	....	....
matching? (Yes / No)	....	....	....	....
Answer to final greeting (spoken)	☐	☐	☐	☐
response time?	....	....	....	....
matching? (Yes / No)	....	....	....	....
Answer to final greeting (gesture)	☐	☐	☐	☐
response time?	....	....	....	....
matching? (Yes / No)	....	....	....	....
Eyebrow frowning	☐	☐	☐	☐
Eyebrow raising	☐	☐	☐	☐
Smiling	☐	☐	☐	☐
Laughing	☐	☐	☐	☐
Lack of interest	☐	☐	☐	☐
Atonishment (mouth open / spoken)	☐	☐	☐	☐
Praise (spoken comment)	☐	☐	☐	☐
Dislike (spoken comment)	☐	☐	☐	☐
Trying to speak	☐	☐	☐	☐
Head shaking	☐	☐	☐	☐
Neck movement	☐	☐	☐	☐
Looking at people around	☐	☐	☐	☐
Other	....	....	....	....

Figure B.4 Checklist used in Egypt and Japan.

B.2 Questionnaire used in Germany

This questionnaire was used after every interaction with ARMAR-IIIb. The context settings in Part 1 were already defined before the experiment.

PART 1						
Your age?	☐					
Your gender?	Male ☐	Female ☐				
Location	Private ☐	Public ☐	Workplace ☐			
Are you familiar with ARMAR?	Yes (co-worker) ☐	Not much (acquaintance) ☐	Met for the first time (unknown) ☐			
Power relationship	Inferior (robot is superior to you) ☐	Equal ☐	Superior (robot is inferior or under your control) ☐			
Time of the day	Morning ☐	Afternoon ☐	Evening ☐			
PART 2	**Please rate ARMAR's way of greeting**					
What was the greeting gesture?	Bow ☐	Nod ☐	Raise hand / wave ☐	Handshake ☐	Hug ☐	
Was it appropriate for the present context?	Inappropriate 1 ☐	2 ☐	3 ☐	4 ☐	5 Appropriate ☐	
What were the greeting words?	Guten Tag ☐	Guten Morgen ☐	Vielen Dank für Ihre Mühe ☐	Hallo! ☐	Guten Abend ☐	Schön dich kennenzulernen ☐
Were they appropriate for the present context?	Inappropriate 1 ☐	2 ☐	3 ☐	4 ☐	5 Appropriate ☐	

In case of gesture score lesser or equal to 3, please fill PART 3
In case of words score lesser or equal to 3, please fill PART 4

Figure B.5 Parts 1 and 2 of the questionnaire used in Germany.

PART 3						
Which greeting type would have been most appropriate in the present context?	Bow ☐	Nod ☐	Raise hand / wave ☐	Handshake ☐	Hug ☐	
Conversely, is there a context in which it the present way of greeting gesture would have been appropriate?	Yes ☐ (please specify below)		No, this greeting is never appropriate ☐ (please skip below)			
Gender (your)	Male ☐	Female ☐	No difference / unrelated ☐			
Location	Private ☐	Public ☐	Workplace ☐	No difference / unrelated ☐		
Social distance	Close ☐	Acquaintance ☐	Unknown (meet the first time) ☐	No difference / unrelated ☐		
Power relationship	Inferior (robot is superior to you) ☐	Equal ☐	Superior (robot is inferior to you) ☐	No difference / unrelated ☐		

PART 4						
Which greeting type would have been most appropriate in the present context?	Guten Tag ☐	Guten Morgen ☐	Vielen Dank für Ihre Mühe ☐	Hallo! ☐	Guten Abend ☐	Schön dich kennenzulernen ☐
Conversely, is there a context in which it the present way of greeting words would have been appropriate?	Yes ☐ (please specify below)		No, this greeting is never appropriate ☐ (please skip below)			
Location	Private ☐	Public ☐	Workplace ☐	No difference / unrelated ☐		
Social distance	Close ☐	Acquaintance ☐	Unknown (meet the first time) ☐	No difference / unrelated ☐		
Power relationship	Inferior (robot is superior to you) ☐	Equal ☐	Superior (robot is inferior to you) ☐	No difference / unrelated ☐		
Time of the day	Morning ☐	Afternoon ☐	Evening ☐	No difference / unrelated ☐		

Figure B.6 Parts 3 and 4 of the questionnaire used in Germany.

References

[1] P. Flandorfer, “Population Ageing and Socially Assistive Robots for Elderly Persons: The Importance of Sociodemographic Factors for User Acceptance”, *International Journal of Population Research*, vol. 2012, Apr. 2012.

[2] A. Weiss and V. Evers, “Exploring cultural factors in human-robot interaction: A matter of personality?”, presented at the 2nd International Workshop on Comparative Informatics, IWCI 2011, Denmark, 2011.

[3] F. Eyssel and D. Kuchenbrandt, “Social categorization of social robots: Anthropomorphism as a function of robot group membership”, *British Journal of Social Psychology*, vol. 51, no. 4, pp. 724–731, 2012.

[4] T. Nomura, T. Suzuki, T. Kanda, J. Han, N. Shin, J. Burke, and K. Kato, “WHAT PEOPLE ASSUME ABOUT HUMANOID AND ANIMAL-TYPE ROBOTS: CROSS-CULTURAL ANALYSIS BETWEEN JAPAN, KOREA, AND THE UNITED STATES”, *International Journal of Humanoid Robotics*, vol. 05, no. 01, pp. 25–46, Mar. 2008.

[5] H. R. Lee, J. Sung, S. Sabanovic, and J. Han, “Cultural design of domestic robots: A study of user expectations in Korea and the United States”, in *2012 IEEE RO-MAN*, 2012, pp. 803–808.

[6] I. Asimov, *The Machine and the Robot. Science Fiction: Contemporaty Mythology*, P.S. Warrick, M. H. Greenberg, J.D. Olander. Harper and Row, 1978.

[7] F. L. Schodt, *Inside the Robot Kingdom - Japan, Mechatronics, and Coming Robotopia*, Kodansha. Tokyo, New York, 1988.

[8] B. Latour, *Nous n'avons jamais été modernes (La découverte). English translation revised and augmented: We Have Never Been Modern*, Simon and Schuster. 31 England and Harvard University Press, 1991.

[9] R. E. Nisbett, *The geography of thought: how Asians and Westerners think differently-- and why*. New York: Free Press, 2004.

[10] F. Kaplan, "WHO IS AFRAID OF THE HUMANOID? INVESTIGATING CULTURAL DIFFERENCES IN THE ACCEPTANCE OF ROBOTS", *International Journal of Humanoid Robotics*, vol. 01, no. 03, pp. 465–480, Sep. 2004.

[11] M. Rosheim, *Leonardo´s Lost Robots*. Springer, 2006.

[12] J. P. Telotte, *Replications: a robotic history of the science fiction film*. University of Illinois Press, 1995.

[14] C. Bartneck, T. Suzuki, T. Kanda, and T. Nomura, "The influence of people's culture and prior experiences with Aibo on their attitude towards robots", *AI & SOCIETY*, vol. 21, no. 1–2, pp. 217–230, May 2006.

[15] C. Bartneck, T. Nomura, T. Kanda, T. Suzuki, and K. Kato, "A cross-cultural study on attitudes towards robots", presented at the Proceedings of the HCI International, 2005.

[16] P. Dario, E. Guglielmelli, C. Laschi, and G. Teti, "MOVAID: a personal robot in everyday life of disabled and elderly people", *Technology and Disability*, vol. 10, no. 2, pp. 77–93, Jan. 1999.

[17] V. Evers, H. C. Maldonado, T. L. Brodecki, and P. J. Hinds, "Relational vs. group self-construal: untangling the role of national culture in HRI", in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, New York, NY, USA, 2008, pp. 255–262.

[18] P. L. P. Rau, Y. Li, and D. Li, "Effects of communication style and culture on ability to accept recommendations from robots", *Comput. Hum. Behav.*, vol. 25, no. 2, pp. 587–595, Mar. 2009.

[19] K. O. Arras and D. Cerqui, "Do we want to share our lives and bodies with robots? A 2000 people survey." Technical Report Nr. 0605-001, Autonomous Systems Lab, Swiss Federal Institute of Technology Lausanne (EPFL), 2005.

[20] M. Makatchev, R. Simmons, M. Sakr, and M. Ziadee, "Expressing Ethnicity through Behaviors of a Robot Character", *arXiv:1303.3592*, Mar. 2013.

[21] L. Riek, N. Mavridis, S. Antali, N. Darmaki, Z. Ahmed, M. Neyadi, and A. Ketheri, "Ibn Sina steps out: exploring Arabic attitudes toward humanoid robots", in *Proceedings of the 2nd international symposium on new frontiers in human–robot interaction, AISB, Leicester*, 2010, vol. 1.

[22] N. Mavridis, M.-S. Katsaiti, S. Naef, A. Falasi, A. Nuaimi, H. Araifi, and A. Kitbi, "Opinions and attitudes toward humanoid robots in the Middle East", *AI & SOCIETY*, vol. 27, no. 4, pp. 517–534, Dec. 2011.

[23] M. Saif, "The Evolution of Persian Thought regarding Art and Figural Representation in Secular and Religious Life after the Coming of Islam", *Macalester Islam Journal*, vol. 1, no. 2, Oct. 2006.

[24] D. W. Straub, K. D. Loch, and C. E. Hill, "Advanced topics in global information management", F. B. Tan, Ed. Hershey, PA, USA: IGI Publishing, 2003, pp. 141–172.

[25] A. Albirini, "Teachers' attitudes toward information and communication technologies: the case of Syrian EFL teachers", *Computers & Education*, vol. 47, no. 4, pp. 373–398, Dec. 2006.

[26] R. M. Thomas, "Computer technology: an example of decision-making in technology transfer", in *Educational Technology—Its Creation, Development and Cross-Cultural Transfer*, R. M. Thomas & V. N. Kobayashi., Oxford: Pergamum Press, pp. 25–34.

[27] E. M. Rogers, *Diffusion of Innovations, 5th Edition*, 5th edition. Free Press, 2003.

[28] M. L. Knapp, *Essentials of Nonverbal Communication*. Harcourt School, 1980.

[29] A. Mehrabian and M. Wiener, "Decoding of inconsistent communications", *Journal of Personality and Social Psychology*, vol. 6, no. 1, pp. 109–114, 1967.

[30] A. Mehrabian and J. T. Friar, "Encoding of attitude by a seated communicator via posture and position cues", *Journal of Consulting and Clinical Psychology*, vol. 33, no. 3, pp. 330–336, 1969.

[31] M. L. Patterson, J. L. Powell, and M. G. Lenihan, “Touch, compliance, and interpersonal affect”, *J Nonverbal Behav*, vol. 10, no. 1, pp. 41–50, Mar. 1986.

[32] A. Mehrabian, *Silent Messages: Implicit Communication of Emotions and Attitudes*. Wadsworth Publishing Company, 1972.

[33] I. Poggi and U. R. Tre, “Signals and meanings of gaze in animated faces”, in *In Language, Vision and Music*, 2001, pp. 133–144.

[34] C. Pelachaud and I. Poggi, “Subtleties of facial expressions in embodied agents”, *The Journal of Visualization and Computer Animation*, vol. 13, no. 5, pp. 301–312, 2002.

[35] M. G. V. Vinayagamoorthy, “Building Expression into Virtual Characters”, *Vinayagamoorthy, V. and Gillies, M. and Steed, A. and Tanguy, E. and Pan, X. and Loscos, C. and Slater, M. (2006) Building Expression into Virtual Characters. In: Eurographics 2006, 04-08 Sept 2006, Vienna, Austria.*, 2006.

[36] M. Argyle, *Bodily Communication*. Methuen, 1988.

[37] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone, “ANIMATED CONVERSATION: Rule-based Generation of Facial Expression, Gesture & Spoken Intonation for Multiple Conversational Agents”, *Technical Reports (CIS)*, May 1994.

[38] S. Duck, *Human Relationships*. SAGE, 2007.

[39] S. R. Fussell, *The Verbal Communication of Emotions: Interdisciplinary Perspectives*. Psychology Press, 2002.

[40] P. Ekman, E. R. Sorenson, and W. V. Friesen, “Pan-cultural elements in facial displays of emotion”, *Science*, vol. 164, no. 3875, pp. 86–88, Apr. 1969.

[41] I. Poggi, *Le parole del corpo. Introduzione alla comunicazione multumodale*. Carocci, 2006.

[42] P. Ekman, *Telling lies: clues to deceit in the marketplace, politics, and marriage*. New York, NY: W.W. Norton, 2009.

[43] S. Marcos, J. Gómez-García-Bermejo, and E. Zalama, "A realistic, virtual head for human–computer interaction", *Interact. Comput.*, vol. 22, no. 3, pp. 176–192, May 2010.

[44] S. Marcos, J. Gomez-Garcia-Bermejo, E. Zalama, and J. Lopez, "Nonverbal communication with a multimodal agent via facial expression recognition", in *2011 IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 1199–1204.

[45] J. Cassell, T. Stocky, T. Bickmore, Y. Gao, Y. Nakano, K. Ryokai, D. Tversky, C. Vaucelle, and H. Vilhjálmsson, "MACK: Media lab Autonomous Conversational Kiosk", in *Proceedings of IMAGINA'02*, Monte Carlo, 2002.

[46] T. W. Bickmore, *Relational Agents: Effecting Change through Human-Computer Relationships*. 2003.

[47] J. Bates, "The role of emotion in believable agents", *Commun. ACM*, vol. 37, no. 7, pp. 122–125, Jul. 1994.

[48] R. G. Harper, A. N. Wiens, and J. D. Matarazzo, *Nonverbal Communication: The State of the Art*. John Wiley & Sons, 1978.

[49] S. Weitz, *Nonverbal communication: readings with commentary*. Oxford University Press, 1974.

[50] H. Aviezer, R. R. Hassin, J. Ryan, C. Grady, J. Susskind, A. Anderson, M. Moscovitch, and S. Bentin, "Angry, disgusted, or afraid? Studies on the malleability of emotion perception", *Psychol Sci*, vol. 19, no. 7, pp. 724–732, Jul. 2008.

[51] R. J. Dolan, J. S. Morris, and B. de Gelder, "Crossmodal binding of fear in voice and face", *PNAS*, vol. 98, no. 17, pp. 10006–10010, Aug. 2001.

[52] G. Pourtois, B. de Gelder, A. Bol, and M. Crommelinck, "Perception of facial expressions and voices and of their combination in the human brain", *Cortex*, vol. 41, no. 1, pp. 49–59, Feb. 2005.

[53] B. de Gelder and J. Van den Stock, "Real Faces, Real Emotions: perceiving Facial Expressions in Naturalistic Contexts of Voices, Bodies, and Scenes", in

Oxford Handbook of Face Perception, G. Rhodes, A. Calder, M. Johnson, and J. V. Haxby, Eds. Oxford University Press, 2011.

[54] H. R. Knudsen and L. H. Muzekari, "The effects of verbal statements of context on facial expressions of emotion", *J Nonverbal Behav*, vol. 7, no. 4, pp. 202–212, Jun. 1983.

[55] D. W. Massaro and P. B. Egan, "Perceiving affect from the voice and the face", *Psychonomic Bulletin & Review*, vol. 3, no. 2, pp. 215–221, Jun. 1996.

[56] R. L. Birdwhistell, *Kinesics and Context: Essays on Body Motion Communication*. University of Pennsylvania Press, 2011.

[57] M. Yuki, W. W. Maddux, and T. Masuda, "Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States", *Journal of Experimental Social Psychology*, vol. 43, no. 2, pp. 303–311, Mar. 2007.

[58] R. Buck, *The communication of emotion*. New York: Guilford Press, 1984.

[59] H. A. Elfenbein and N. Ambady, "Universals and Cultural Differences in Recognizing Emotions", *Current Directions in Psychological Science*, vol. 12, no. 5, pp. 159–164, Oct. 2003.

[60] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion", *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.

[61] K. Shimoda, M. Argyle, and P. R. Bitti, "The intercultural recognition of emotional expressions by three national racial groups: English, Italian and Japanese", *European Journal of Social Psychology*, vol. 8, no. 2, pp. 169–179, 1978.

[62] T. Koda, M. Rehm, and E. André, *Cross-cultural Evaluations of avatar facial expressions designed by Western Designers*. 2008.

[63] T. Koda, "Cross-Cultural Study of Avatars' Facial Expressions and Design Considerations Within Asian Countries", in *Intercultural Collaboration*, T. Ishida, S. R. Fussell, and P. T. J. M. Vossen, Eds. Springer Berlin Heidelberg, 2007, pp. 207–220.

[64] R. Beira, M. Lopes, M. Praga, J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi, and R. Saltaren, “Design of the robot-cub (iCub) head”, in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*, 2006, pp. 94–100.

[65] J. Oh, D. Hanson, W. Kim, Y. Han, J. Kim, and I. Park, “Design of Android type Humanoid Robot Albert HUBO”, 2006, pp. 1428–1433.

[66] K. Itoh, H. Miwa, M. Matsumoto, M. Zecca, H. Takanobu, S. Roccella, M. C. Carrozza, P. Dario, and A. Takanishi, “Various emotional expressions with emotion expression humanoid robot WE-4RII”, in *First IEEE Technical Exhibition Based Conference on Robotics and Automation, TExCRA ’04*, 2004, pp. 35–36.

[67] N. Endo and A. Takanishi, “Development of Whole-Body Emotional Expression Humanoid Robot for ADL-Assistive RT Services”, *Journal of Robotics and Mechatronics*, vol. 23, no. 6, pp. 969–977, 2011.

[68] C. Becker-Asano and H. Ishiguro, “Evaluating facial displays of emotion for the android robot Geminoid F”, in *2011 IEEE Workshop on Affective Computational Intelligence (WACI)*, 2011, pp. 1–8.

[69] F. Hegel, F. A. Eyssel, and B. Wrede, “The Social Robot Flobi: Key Concepts of Industrial Design”, in *Proceedings of the 19th IEEE International Symposium in Robot and Human Interactive Communication (RO-MAN 2010)*, 2010, pp. 120–125.

[70] A. van Breemen, X. Yan, and B. Meerbeek, “iCat: an animated user-interface robot with personality”, in *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, New York, NY, USA, 2005, pp. 143–144.

[71] J. Kędzierski, R. Muszyński, C. Zoll, A. Oleksy, and M. Frontkiewicz, “EMYS—Emotive Head of a Social Robot”, *Int J of Soc Robotics*, vol. 5, no. 2, pp. 237–249, Apr. 2013.

[72] C. Breazeal, “Emotion and sociable humanoid robots”, *Int. J. Hum.-Comput. Stud.*, vol. 59, no. 1–2, pp. 119–155, Jul. 2003.

[73] D. Mazzei, N. Lazzeri, D. Hanson, and D. De-Rossi, “HEFES: An Hybrid Engine for Facial Expressions Synthesis to control human-like androids and avatars”, in *2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2012, pp. 195–200.

[74] A. Beck, A. Hiolle, A. Mazel, and L. Cañamero, “Interpretation of emotional body language displayed by robots”, in *Proceedings of the 3rd international workshop on Affective interaction in natural environments*, New York, NY, USA, 2010, pp. 37–42.

[75] I. Poggi, “From a Typology of Gestures to a Procedure for Gesture Production”, in *Gesture and Sign Language in Human-Computer Interaction*, I. Wachsmuth and T. Sowa, Eds. Springer Berlin Heidelberg, 2002, pp. 158–168.

[76] D. Morris, *Bodytalk: a world guide to gestures*. Jonathan Cape, 1994.

[77] S. Hoffman-Hicks, “The longitudinal development of French foreign language pragmatic competence: Evidence from study abroad participants.”, Doctoral Dissertation, Indiana University, 1999.

[78] A. Duranti, “Universal and Culture-Specific Properties of Greetings”, *Journal of Linguistic Anthropology*, vol. 7, no. 1, pp. 63–97, 1997.

[79] S. K. Fan, “Japanese communication beyond language : From the viewpoint of Hong Kong Chinese”, *The bulletin of the Institute for Japan Studies, Kanda University of International Studies*, vol. 3, p. 35_a–21_a, Mar. 2002.

[80] Y. H. Zoubir, “Doing Business in Egypt”, *Thunderbird International Business Review*, vol. 42, no. 3, pp. 329–347, 2000.

[81] J. Huwer, “The effect of context, intimacy and gender on Handshaking”, Senior Thesis Project, Department of Psychology, Haverford College, 2003.

[82] K. Nishiyama, *Doing Business With Japan: Successful Strategies for Intercultural Communication*. University of Hawaii Press, 2000.

[83] H. S. Friedman, R. E. Riggio, and M. R. Di Matteo, “A classification of nonverbal greetings for use in studying face-to-face interaction”, *JSAS Catalog of Selected Documents in Psychology*, vol. 11, pp. 31–32, 1981.

[84] D. P. E. Greenbaum and H. M. Rosenfeld, “Varieties of touching in greetings: Sequential structure and sex-related differences”, *J Nonverbal Behav*, vol. 5, no. 1, pp. 13–25, Sep. 1980.

[85] S. Sugito, *Aisatsu no kotoba to miburi (あいさつの言葉と身振り)*. Bunkachou, 1981.

[86] H. Spencer-Oatey, “Reconsidering power and distance”, *Journal of Pragmatics*, vol. 26, no. 1, pp. 1–24, Jul. 1996.

[87] R. Firth, “Verbal and bodily rituals of greeting and parting”, in *The Interpretation of Ritual: Essays in Honour of A.I. Richards*, J. La Fontaine, Ed. Tavistock Publications, 1972, pp. 1–38.

[88] S. E. Scherer and M. R. Schiff, “Perceived intimacy, physical distance and eye contact”, *Percept Mot Skills*, vol. 36, no. 3, pp. 835–841, Jun. 1973.

[89] T. Nguyen, R. Heslin, and M. L. Nguyen, “The Meanings of Touch: Sex Differences”, *Journal of Communication*, vol. 25, no. 3, pp. 92–103, 1975.

[90] R. E. Riggio, H. S. Friedman, and M. R. DiMatteo, “Nonverbal Greetings: Effects of the Situation and Personality”, *Personality and Social Psychology Bulletin*, vol. 7, no. 4, pp. 682–689, Dec. 1981.

[91] T. C. Marshall, “Cultural differences in intimacy: The influence of gender-role ideology and individualism—collectivism”, *Journal of Social and Personal Relationships*, vol. 25, no. 1, pp. 143–168, Feb. 2008.

[92] P. Brown and S. C. Levinson, *Politeness: Some Universals in Language Usage*. Cambridge University Press, 1987.

[93] L. Hickey and M. Stewart, *Politeness in Europe*. Multilingual Matters, 2005.

[94] B. R. Slugoski and W. Turnbull, “Cruel to be Kind and Kind to be Cruel: Sarcasm, Banter and Social Relations”, *Journal of Language and Social Psychology*, vol. 7, no. 2, pp. 101–121, Jun. 1988.

[95] C. A. Ferguson, “The structure and use of politeness formulas”, *Language in Society*, vol. 5, no. 02, pp. 137–151, 1976.

[96] C. Kern and E. Eichmüller, *Politeness strategies in Hungary and England with special focus on greetings and leave-taking terms*, 1 edition. GRIN Verlag GmbH, 2007.

[97] D. Li, "A Contrastive Study Of Greetings And Partings In English And Chinese", Department of Foreign Languages, Guangdong Women's Professional Technical College, Guangzhou, China.

[98] T. Asfour, K. Regenstein, P. Azad, J. Schroder, A. Bierbaum, N. Vahrenkamp, and R. Dillmann, "ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control", in *2006 6th IEEE-RAS International Conference on Humanoid Robots*, 2006, pp. 169–175.

[99] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, "The intelligent ASIMO: system overview and integration", in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2002*, 2002, vol. 3, pp. 2478–2483 vol.3.

[100] K. Kaneko, F. Kanehiro, M. Morisawa, K. Akachi, G. Miyamori, A. Hayashi, and N. Kanehira, "Humanoid robot HRP-4 - Humanoid robotics platform with lightweight and slim body", in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2011, pp. 4400–4407.

[101] Y. Bum-Jae, "Network-Based Humanoids 'MAHRU' As Ubiquitous Robotic Companion", 2008, pp. 724–729.

[102] M. Yamamoto and T. Watanabe, "Time delay effects of utterance to communicative actions on greeting interaction by using a voice-driven embodied interaction system", in *2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation, 2003. Proceedings*, 2003, vol. 1, pp. 217–222 vol.1.

[103] S. Suzuki, Y. Fujimoto, and T. Yamaguchi, "Can differences of nationalities be induced and measured by robot gesture communication ?", in *2011 4th International Conference on Human System Interactions (HSI)*, 2011, pp. 357–362.

[104] H. Aoki, Y. Fujimoto, S. Suzuki, E. Sato-Shimokawara, and T. Yamaguchi, "Interaction robot considered difference of culture and physiological

responses on greeting with the robot", in *2011 Proceedings of SICE Annual Conference (SICE)*, 2011, pp. 2510–2513.

[105] N. L. Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*. Fortress Press, 2002.

[106] T. Kishi, T. Otani, N. Endo, P. Kryczka, K. Hashimoto, K. Nakata, and A. Takanishi, "Development of expressive robotic head for bipedal humanoid robot", presented at the Intelligent Robots and Systems (IROS), 2012, pp. 4584–4589.

[107] Y. Ogura, H. Aikawa, H. Kondo, A. Morishima, H. Lim, and A. Takanishi, "Development of a new humanoid robot WABIAN-2", in *in Proceedings of IEEE International Conference on Robotics and Automation*, 2006, pp. 76–81.

[108] A. Albers, S. Brudniok, J. Ottnad, C. Sauter, and K. Sedchaicharn, "Upper Body of a new Humanoid Robot - the Design of ARMAR III", in *2006 6th IEEE-RAS International Conference on Humanoid Robots*, 2006, pp. 308–313.

[109] T. Asfour, K. Berns, and R. Dillmann, "The humanoid robot ARMAR: Design and control", in *IN IEEE/APS INTL CONFERENCE ON HUMANOID ROBOTS*, 2000, pp. 7–8.

[110] R. Dillmann, R. Becher, and P. Steinhaus, "ARMAR II — A LEARNING AND COOPERATIVE MULTIMODAL HUMANOID ROBOT SYSTEM", *International Journal of Humanoid Robotics*, vol. 01, no. 01, pp. 143–155, Mar. 2004.

[111] G. Trovato, T. Kishi, N. Endo, K. Hashimoto, and A. Takanishi, "Development of Facial Expressions Generator for Emotion Expressive Humanoid Robot", in *Proceedings of 2012 IEEE-RAS International Conference on Humanoid Robots*, Osaka, Japan.

[112] G. Trovato, M. Zecca, T. Kishi, N. Endo, K. Hashimoto, and A. Takanishi, "GENERATION OF HUMANOID ROBOT'S FACIAL EXPRESSIONS FOR CONTEXT-AWARE COMMUNICATION", *International Journal of Humanoid Robotics*, vol. 10, no. 01, p. 1350013, 2013.

[113] I. Poggi and C. Pelachaud, "Performative Facial Expressions in Animated Faces", in *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, E. Churchill., Cambridge: MIT press, 2000.

[114] J. A. Russell, “A circumplex model of affect”, *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[115] J. A. Russell, M. Lewicka, and T. Niit, “A cross-cultural study of a circumplex model of affect”, *Journal of Personality and Social Psychology*, vol. 57, no. 5, pp. 848–856, 1989.

[116] A. Mehrabian, *Basic dimensions for a general psychological theory: implications for personality, social, environmental, and developmental studies*. Oelgeschlager, Gunn & Hain, 1980.

[117] J. R. J. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, “The world of emotions is not two-dimensional”, *Psychol Sci*, vol. 18, no. 12, pp. 1050–1057, Dec. 2007.

[118] J. Ahn, S. Gobron, Q. Silvestre, and D. Thalmann, “Asymmetrical Facial Expressions based on an Advanced Interpretation of Two-dimensional Russell’s Emotional Model”, in *Proceedings of ENGAGE*, 2010.

[119] K. R. Scherer, “What are emotions? And how can they be measured?”, *Social Science Information*, vol. 44, no. 4, pp. 695–729, Dec. 2005.

[120] K. R. Scherer, V. Shuman, J. R. J. Fontaine, and C. Soriano, “The GRID meets the Wheel: Assessing emotional feeling via self-report”, in *Components of emotional meaning: A sourcebook*, Oxford: Oxford University Press.

[121] H. Miwa, T. Okuchi, H. Takanobu, and A. Takanishi, “Development of a new human-like head robot WE-4”, in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2002*, 2002, vol. 3, pp. 2443–2448 vol.3.

[122] R. Plutchik, *Emotions and Life: Perspectives from Psychology, Biology, and Evolution*. Washington, DC: American Psychological Association, 2002.

[123] E. Douglas-Cowie, J. C. Martin, J. Devillers, and C. Cox, “HUMAINE project, Mid Term Report on Database Exemplar Progress.”

[124] P. Ekman and W. V. Friesen, *Facial Action Coding System*. Consulting Psychologists Press, 1978.

[125] J. C. Hager and P. Ekman, "The Inner And Outer Meanings Of Facial Expressions", in *Social Psychophysiology: A Sourcebook*, J. T. Cacioppo & R. E. Petty., New York: The Guilford Press, 1983.

[126] N. Stoiber, R. Seguier, and G. Breton, "Facial animation retargeting and control based on a human appearance space", *Computer Animation and Virtual Worlds*, vol. 21, no. 1, pp. 39–54, 2010.

[127] J. Saldien, K. Goris, B. Vanderborght, J. Vanderfaeillie, and D. Lefeber, "Expressing Emotions with the Social Robot Probo", *Int J of Soc Robotics*, vol. 2, no. 4, pp. 377–389, Dec. 2010.

[128] T. Ribeiro and A. Paiva, "The illusion of robotic life: principles and practices of animation for robots", in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, New York, NY, USA, 2012, pp. 383–390.

[129] I. Poggi, "The lexicon of the conductor's face", in *Proceedings of the Workshop on "Language, vision and music",* 1999, pp. 67–54.

[130] I. Poggi, "Towards the Alphabet and the Lexicon of Gesture, Gaze and Touch", presented at the Multimodality of Human Communication. Theories, problems and applications, Virtual Symposium, 2001.

[131] H. Kobayashi, Y. Ichikawa, M. Senda, and T. Shiiba, "Toward rich facial expression by face robot", in *Proceedings of 2002 International Symposium on Micromechatronics and Human Science, 2002. MHS 2002*, 2002, pp. 139–145.

[132] C. A. Smith and H. S. Scott, "A Componential Approach to the meaning of facial expressions", in *The psychology of facial expression*, J. A. Russell and J. M. Fern, Eds. Paris, France: Editions de la Maison des Sciences de l'Homme, 1997, pp. 229–254.

[133] D. L. Bimler and G. V. Paramei, "Facial-Expression Affective Attributes and their Configural Correlates: Components and Categories", *The Spanish Journal of Psychology*, vol. 9, no. 1, pp. 19 – 31, Apr. 2006.

[134] S. Raudys and R. P. W. Duin, "Expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix", *Pattern Recogn. Lett.*, vol. 19, no. 5–6, pp. 385–392, Apr. 1998.

[135] G. Trovato, T. Kishi, N. Endo, K. Hashimoto, and A. Takanishi, “Evaluation study on asymmetrical facial expressions generation for Humanoid Robot”, in *2012 First International Conference on Innovative Engineering Systems (ICIES)*, Alexandria, Egypt, 2012, pp. 129–134.

[136] G. Trovato, T. Kishi, N. Endo, M. Zecca, K. Hashimoto, and A. Takanishi, “Cross-Cultural Perspectives on Emotion Expressive Humanoid Robotic Head: Recognition of Facial Expressions and Symbols”, *Int J of Soc Robotics,*. vol. 5, no. 4, pp. 515–527, 2013.

[137] G. Trovato, T. Kishi, N. Endo, K. Hashimoto, and A. Takanishi, “A Cross-Cultural Study on Generation of Culture Dependent Facial Expressions of Humanoid Social Robot”, presented at the International Conference on Social Robotics (ICSR 2012), Chengdu, China, pp. 35–44.

[138] F. Thomas and O. Johnston, *The Illusion of Life: Disney Animation*, Rev Sub. Disney Editions, 1995.

[139] J. Ahn, S. Gobron, Q. Silvestre, D. Thalmann, and R. Boulic, “Double your Face with Asymmetric Expression of Emotions Driven by a 3D Emotional Model”, presented at the Eurographics/ACM SIGGRAPH Symposium on Computer Animation, 2011.

[140] J. C. Borod, C. S. Haywood, and E. Koff, “Neuropsychological aspects of facial asymmetry during emotional expression: a review of the normal adult literature”, *Neuropsychol Rev*, vol. 7, no. 1, pp. 41–60, Mar. 1997.

[141] J. C. Borod, E. Koff, S. Yecker, C. Santschi, and J. M. Schmidt, “Facial asymmetry during emotional expression: gender, valence, and measurement technique”, *Neuropsychologia*, vol. 36, no. 11, pp. 1209–1215, Nov. 1998.

[142] K. Grammer and R. Thornhill, “Human (Homo sapiens) facial attractiveness and sexual selection: The role of symmetry and averageness.”, *Journal of Comparative Psychology*, vol. 108, no. 3, pp. 233–242, 1994.

[143] K. L. Schmidt, Y. Liu, and J. F. Cohn, “The role of structural facial asymmetry in asymmetry of peak facial expressions”, *Laterality*, vol. 11, no. 6, pp. 540–561, Nov. 2006.

[144] R. S. van Gelder and J. C. Borod, “Neurobiological and cultural aspects of facial asymmetry”, *J Commun Disord*, vol. 23, no. 4–5, pp. 273–286, Oct. 1990.

[145] H. S. C. J C Borod, “Asymmetry in positive and negative facial expressions: sex differences.”, *Neuropsychologia*, vol. 19, no. 6, pp. 819–24, 1981.

[146] A. Hosoi, *Kyarakutaa no yukata na kanjou hyougen no kakikata*. Seibundo-shinkosha, Tokyo.

[147] M. Argyle and P. Trower, *Person to Person: Ways of Communicating*. Harper & Row., 1979.

[148] E. T. Hall, *Beyond Culture*. Anchor Books, 1976.

[149] D. Li, P. L. P. Rau, and Y. Li, “A Cross-cultural Study: Effect of Robot Appearance and Task”, *Int J of Soc Robotics*, vol. 2, no. 2, pp. 175–186, Jun. 2010.

[150] G. Trovato, M. Zecca, S. Sessa, L. Jamone, J. Ham, K. Hashimoto, and A. Takanishi, “Towards culture-specific robot customisation: a study on greeting interaction with Egyptians”, presented at the The 22nd IEEE International Symposium on Robot and Human Interactive Communication (ROMAN 2013), Gueongju, South Korea, 2013.

[151] G. Trovato, M. Zecca, S. Sessa, L. Jamone, J. Ham, K. Hashimoto, and A. Takanishi, “Cross-cultural study on human-robot greeting interaction: acceptance and discomfort by Egyptians and Japanese”, *Paladyn. International Journal of Behavioral Robotics*, vol. 4, no. 2, pp. 83–93, 2013.

[152] G. Metta, P. Fitzpatrick, and L. Natale, “YARP: Yet Another Robot Platform”, *International Journal of Advanced Robotic Systems*, p. 1, 2006.

[153] T. D. Blumenthal, B. N. Cuthbert, D. L. Filion, S. Hackley, O. V. Lipp, and A. van Boxtel, “Committee report: Guidelines for human startle eyeblink electromyographic studies”, *Psychophysiology*, vol. 42, no. 1, pp. 1–15, Jan. 2005.

[154] P. Cipresso, S. Serino, D. Villani, C. Repetto, L. Sellitti, G. Albani, A. Mauro, A. Gaggioli, and G. Riva, “Is your phone so smart to affect your state? An exploratory study based on psychophysiological measures”, *Neurocomput.*, vol. 84, pp. 23–30, May 2012.

[155] S. Kaiser and T. Wehrle, “Facial expressions as indicators of appraisal processes”, in *Appraisal processes in emotions: Theory, methods, research*, K. R. Scherer, A. Schorr, & T. Johnstone., New York: Oxford University Press, pp. 285–300.

[156] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, “Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots”, *Int J of Soc Robotics*, vol. 1, no. 1, pp. 71–81, Jan. 2009.

[157] W. H. Kruskal and W. A. Wallis, “Use of Ranks in One-Criterion Variance Analysis”, *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.

[158] H. B. Mann and D. R. Whitney, “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other”, *Ann. Math. Statist.*, vol. 18, no. 1, pp. 50–60, Mar. 1947.

[159] D. George and P. Mallery, *SPSS for Windows step by step: a simple guide and reference, 11.0 update*. Boston: Allyn and Bacon, 2003.

[160] B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. The Center for the Study of Language and Information Publications, 1998.

[161] C. Nass and K. M. Lee, “Does computer-synthesized speech manifest personality? experimental tests of recognition, similarity-attraction, and consistency-attraction”, *Journal of Experimental Psychology: Applied*, vol. 7, pp. 171–181, 2001.

[162] C. Nass, Y. Moon, and P. Carney, “Are People Polite to Computers? Responses to Computer-Based Interviewing Systems1”, *Journal of Applied Social Psychology*, vol. 29, no. 5, pp. 1093–1109, 1999.

[163] C. H. Lovelock and L. Wright, *Principles of service marketing and management*. Prentice Hall, 2002.

[164] V. Evers, “Cross-cultural Differences in Understanding Human-Computer Interfaces”, in *HCI ’98 Conference Companion*, Sheffield, UK, 1998.

[165] O. Mizutani and N. Mizutani, *Nihongo notes 1: Speaking and living in Japan*, The Japan Times. Tokyo, 1988.

[166] P. Azad, T. Asfour, and R. Dillmann, "Toward an Unified Representation for Imitation of Human Motion on Humanoids", in *2007 IEEE International Conference on Robotics and Automation*, 2007, pp. 2558–2563.

[167] T. S. Polzin and A. H. Waibel, *Detecting Emotions in Speech*. 1998.

[168] M. Kuramochi, "出会いのあいさつ言葉におけるコードと機能の変化", *Gengo to Kouryu* （言語と交流）, no. 13, 2010.

[169] M. Kuramochi, "あいさつ言葉の生成と変容", Meikai University, 2013.

[170] M. Kuramochi, "Generation and Transformation of Aisatsu Words", Meikai University, 2013.

[171] Lundmark, *Tales of Hi and Bye*. Cambridge University Press.

[172] G. Scagliotti and B. Mujtaba, "Take a bow: culturally preparing expatriates for doing business in Japan", *Journal of Comprehensive Research*, pp. 56–72.

[173] B. D. Mente, *Kata: The Key to Understanding & Dealing with the Japanese!* Tuttle Publishing, 2003.

[174] M. Do, P. Azad, T. Asfour, and R. Dillmann, "Imitation of human motion on a humanoid robot using non-linear optimization", in *8th IEEE-RAS International Conference on Humanoid Robots, 2008. Humanoids 2008*, 2008, pp. 545–552.

[175] T. Kasuga and M. Hashimoto, "Human-Robot Handshaking using Neural Oscillators", in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005. ICRA 2005*, 2005, pp. 3802–3807.

[176] M. Jindai, T. Watanabe, S. Shibata, and T. Yamamoto, "Development of a Handshake Robot System for Embodied Interaction with Humans", in *The 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006. ROMAN 2006*, 2006, pp. 710–715.

[177] S. Mitchell, M. Destephe, G. Trovato, M. Zecca, K. Hashimoto, and A. Takanishi, "Design Choices in the Development of Robotic Heads: Human-

likeness, Form and Colours", presented at the Workshop on "Towards Social Humanoid Robots: What makes interaction human-like?", in RSJ IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, 2013.

[178] K. Petersen, K. Fukui, Z. Lin, N. Endo, E. Kazuki, H. Ishii, M. Zecca, A. Takanishi, T. Asfour, and R. Dillmann, "Towards high-level, cloud-distributed robotic telepresence: Concept introduction and preliminary experiments", in *2011 IEEE RO-MAN*, 2011, pp. 131–136.

Research achievements

International journals

- G. Trovato, M. Zecca, T. Kishi, N. Endo, K. Hashimoto, and A. Takanishi, "Generation of humanoid robot's facial expressions for context-aware communication" *International Journal of Humanoid Robotics*, vol. 10, no. 01, p. 1350013, Mar 2013.
- G. Trovato, T. Kishi, N. Endo, M. Zecca, K. Hashimoto, and A. Takanishi, "Cross-Cultural Perspectives on Emotion Expressive Humanoid Robotic Head: Recognition of Facial Expressions and Symbols" *International Journal of Social Robotics*, vol. 5, no. 4, pp. 515–527, Sep 2013. DOI: 10.1007/s12369-013-0213-z
- G. Trovato, M. Zecca, S. Sessa, L. Jamone, J. Ham, K. Hashimoto, and A. Takanishi, "Cross-cultural study on human-robot greeting interaction: acceptance and discomfort by Egyptians and Japanese" *Paladyn. International Journal of Behavioral Robotics,* vol. 4, no. 2, pp. 83–93, Oct 2013. DOI: 10.2478/pjbr-2013-0006 JBR

International conferences

- G. Trovato, T. Kishi, N. Endo, K. Hashimoto, and A. Takanishi, "A Cross-Cultural Study on Generation of Culture Dependent Facial Expressions of

Humanoid Social Robot", presented at the *International Conference on Social Robotics (ICSR2012)*, Chengdu, China, pp. 35–44, Oct 2012.

- G. Trovato, T. Kishi, N. Endo, K. Hashimoto, and A. Takanishi, "Development of Facial Expressions Generator for Emotion Expressive Humanoid Robot", in *Proceedings of 2012 IEEE-RAS International Conference on Humanoid Robots*, Osaka, Japan, Nov 2012.
- G. Trovato, T. Kishi, N. Endo, K. Hashimoto, and A. Takanishi, "Evaluation study on asymmetrical facial expressions generation for Humanoid Robot", in *2012 First International Conference on Innovative Engineering Systems (ICIES)*, Alexandria, Egypt, pp. 129–134, Dec 2012.
- T. Kishi, T. Kojima, N. Endo, M. Destephe, T. Otani, L. Jamone, P. Kryczka, G. Trovato, K. Hashimoto, S. Cosentino, and A. Takanishi, "Impression Survey of the Emotion Expression Humanoid Robot with Mental Model Based Dynamic Emotions", presented at the *2013 IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, May 2013.
- G. Trovato, M. Zecca, S. Sessa, L. Jamone, J. Ham, K. Hashimoto, and A. Takanishi, "Towards culture-specific robot customisation: a study on greeting interaction with Egyptians" presented at the *The 22nd IEEE International Symposium on Robot and Human Interactive Communication (ROMAN 2013)*, Gyeongju, South Korea, Aug 2013.

Domestic conferences

- G. Trovato, "Generation of Humanoid Robots' Facial Expressions for Intercultural Communication", presented at the *International Symposium on Micro-Nano Systems for the Interaction of Young Researchers*, Nagoya, Japan, Nov 2012.
- H. Futaki, T. Kishi, N. Endo, G. Trovato, T. Otani, A. Iizuka, M. Destephe, K. Hashimoto, and A. Takanishi, "漫画表現に特化した表情表出能力を有する２足ヒューマノイドロボット頭部の開発", presented at the 第 31 回日本ロボット学会学術講演予稿 集 (RSJ), Sep 2013.

Invited lectures

- "Towards robotics: secondary school and interdisciplinary education. Humanoid robotics in Japan and Germany" (in Italian), Scientific Secondary School "F. Cecioni", Livorno, Italy, May 2012
- "A cross-cultural study on humanoid robot facial expressions and symbols display", Beijing Institute of Technology, Beijing, China, Oct 2012.
- "Exploring the psychology of face-to-face communication with the humanoid robot KOBIAN-R", Eindhoven University of Technology, Eindhoven, Netherlands, Jul 2013
- "Towards culture-specific robots: customising the humanoid robot KOBIAN", Korea Advanced Institute of Science & Technology, Daejeon, South Korea, Aug 2013
- "Acceptance and discomfort in human robot interaction through different cultures", Center for Intelligent Robotics, Korea Institute of Science and Technology, Seoul, South Korea, Aug 2013
- "Face-to-face communication and greeting interaction with the humanoid robot KOBIAN", Dongseo University, Busan, South Korea, Aug 2013
- "Design choices and behaviour customisation of robots for different cultures", Bielefeld University, Bielefeld, Germany, Dec 2013
- "Hello human! – Greeting gestures in human-robot interaction", Tilburg University, Tilburg, Netherlands, Jan 2014
- "Customising culture-dependent robot design and non-verbal communication", University of Hertfordshire, Hatfield, UK, Jan 2014
- "Development of culturally adaptive non-verbal communication capabilities for humanoid robots" (in Japanese), Global Robot Academia Symposium 卓越成果報告会, Tokyo, Japan, Feb 2014

Awards

E-JUST

Award Certificate

Best Paper Finalist

ICIES2012

1st International Conference on Innovative Engineering Systems

This is to certify that the paper entitled

" Evluation Study on Asymmetrical Facial Expressions Generation for Humanoid Robot "

by Gabriele Trovato, Tatsuhiro Kishi, Nobutsuna Endo, Kenji Hashimoto, Atsuo Takanishi

Is selected as the best paper finalist in

" Innovative Engineering Systems Conference, ICIES 2012, IEEE/RAS ".

Held in Alexandria, Egypt, 7-9 December, 2012.

IEEE

SICE

Program Chair,

Dr. Mohamed Abdellatif

General Chair,

Prof. Ahmed B. Khairy