

A novel culture-dependent gesture selection system for a humanoid robot performing greeting interaction

G. Trovato¹, M. Do², M. Kuramochi³, M. Zecca⁴, Ö. Terlemez²,
T. Asfour², and A. Takanishi^{5,6}

¹ Faculty of Science and Engineering, Waseda University, #41-304, 17 Kikui-cho,
Shinjuku-ku, Tokyo 162-0044, Japan

² Institute of Anthropomatics, Karlsruhe Institute of Technology, Karlsruhe, Germany

³ Kanda Institute of Foreign Languages, Tokyo, Japan

⁴ School of Electronic, Electrical and Systems Engineering, Loughborough University, UK

⁵ Department of Modern Mechanical Engineering, Waseda University

⁶ Humanoid Robotics Institute (HRI), Waseda University

contact@takanishi.mech.waseda.ac.jp

Abstract. In human-robot interaction, it is important for the robots to adapt to our ways of communication. As humans, rules of non-verbal communication, including greetings, change depending on our culture. Social robots should adapt to these specific differences in order to communicate effectively, as a correct way of approaching often results into better acceptance of the robot. In this study, a novel greeting gesture selection system is presented and an experiment is run using the robot ARMAR-IIIb. The robot performs greeting gestures appropriate to Japanese culture; after interacting with German participants, the selection should become appropriate to German culture. Results show that the mapping of gesture selection evolves successfully.

Keywords: Social Robotics, Culture, Gestures, Greetings, HRI

1 Introduction

The relationship between acceptance of robots into human societies and its background culture is an idea that has been debated since long time. According to the traditional view in literature, anxiety towards robots is more common in Western countries. As a matter of fact, differences between East and West in cognition, due to differing social structures, philosophies, and educational systems, trace back to ancient Greece and China [1]. Stereotypes are not always true, as there are positive examples of robotic heroes in Western science fiction too; however, technology acceptance, for instance, depends also on the country that is the producer, since the culture of that country may bias some aspects of the product. As a consequence, localisation of products may be done [2]. In our previous research [3], a comparative study carried out with Egyptian and Japanese participants, culture-dependent acceptance and discomfort were found relating to greeting gestures of a humanoid robot. As the impor-

tance of culture-specific customisation for acceptance of robots was confirmed, the need of a system of greeting selection for robots was highlighted.

1.1 Greeting interaction and related works

Greeting is the basic way of initiating and closing an interaction. Hoffman-Hicks [4] 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. Moreover, 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, chances of misunderstanding and difficulty in communication. It is then necessary to understand from sociology studies which factors influence greetings between humans. A unified model for greeting does not exist; therefore the study has to be done through a survey of different sources from different countries.

Intimacy and Politeness are two important keywords in sociology, and both influence the choice of a greeting gesture [5, 6]. Intimacy is apparently influenced by Physical Distance, Eye contact [7], Gender [8], Location [5] and Culture [9]. Brown and Levinson [10] were the first to think of a formula for calculating Politeness. Even though they did not define numerically any coefficient, they represented Politeness as a function of Power Relationship, Social Distance and a cultural factor.

Some humanoid robots can perform programmed greetings. Among others is ARMAR-III [11], which met the German Prime Minister Angela Merkel. ASIMO [12] is capable of performing a wider range of greetings: handshake, wave both hands, and bow, and can recognise such gestures among others. MAHRU [13] is another example of humanoid robot which can greet through a simple bow.

While greeting gestures have been programmed, to the best of our knowledge, so far only a few greeting interaction experiments with robots have been conducted to test the impression on humans. Experiments done in [14], which focused on timing, rather than on culture; and experiments featuring the social robot ApriPoco, in which data from biological signals of subjects looking at Japanese, Chinese, and French greetings were compared [15]. Our intention in this experiment is instead to make the robot choose the right greeting, rather than assess human reaction. Another important difference is that our studies were done with a human sized humanoid social robot.

1.2 Objectives of this paper

The main idea behind this study is a typical scenario in which a foreigner in a country visited for the first time (e.g. a Westerner in Japan) greets local people in an 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. However, in a limited number of interactions, the foreigner can understand the rules and correct his behaviour. In the current experiment, we want a robot to be able to do the same: be trained with sociology data related to one country, and evolve its behaviour engaging with people of another country in a small number of interactions. For

the implementation of the gestures and the interaction experiment, we used the humanoid robot ARMAR-IIIb [11].

As the experiment is carried out in Germany, the interactions are done with German participants, while preliminary training is done with Japanese data, which is culturally extremely different. Participants' feedback is also collected, but it is not the main goal of this research, because as previously stated, culture-dependent acceptance and discomfort were already found in the previous Egyptian-Japanese study [3]. The point of interest in this paper is about the evolution of behaviour itself, from Japanese to German.

The rest of the paper is organized as follows: in Section 2 we describe the system of greeting selection; in Section 3 the hardware implementation; in Section 4 we describe the experiment and show and discuss the results; in Section 5 we conclude the paper and outline future works.

2 Greeting selection

2.1 Model of greetings

It is necessary to identify the main factors that influence the choice of gesture in human greetings. In Section 1.1 the main factors are reported, and following a process of simplification, the resulting factors are summarised in Figure 1. Simplification consisted in dropping some factors, such as physical distance and eye contact, assuming that they are always fixed or guaranteed. The remaining factors (Culture, Gender, Location, Power relationship, Social distance) are therefore relative to only social aspects of interaction. They are listed on the left of Figure 1 with their possible values. These factors are the features of the problem of mapping an input containing this social context information into a greeting gesture selection on the right of Figure 1. The possible values of the features are categorical data, as they can assume only 2 or 3 values, and are given as input to the mapping problem. Culture is a special case as it can be considered a discriminant for switching among different mappings between the other factors and the outputs.

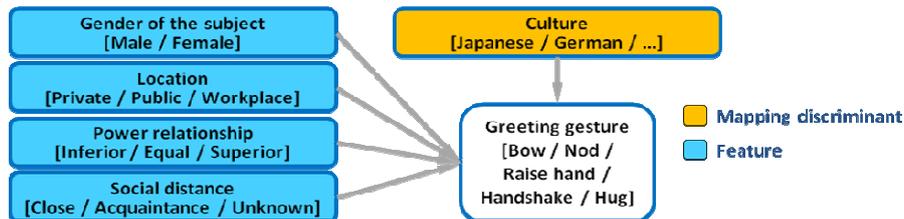


Figure 1. The model of greeting selection synthesised in four features on the left, one mapping discriminant and the output on the right. Each block has some possible categorical values.

The outputs can also assume only a limited set of categorical values, the classes of a mapping problem. Greeting gestures list (Bow, Nod, Handshake, Raise hand, Hug)

has been defined from relevant sources [5], [16, 17]. Originally, the set contained six gesture types, including kissing, which was dropped, because it was not possible to implement on the robot ARMAR-IIIb, which does not have a mouth. Waving and raising a hand or two hands were also considered as broadly the same type of gesture.

2.2 Greeting selection system

Figure 2 contains the overview of the greeting selection system. It takes context data (Gender, Location, and so on) as input and produces the appropriate robot posture (the configuration for the chosen gesture) for that input. The context is the set of features shown in Figure 1. Inside the mapping box there is an algorithm that will be described in Section 2.3. The gesture chosen from this mapping is turned into robot configuration through the Master Motor Map (in short MMM) [18], which will be described more in detail in Section 3. The output gets evaluated by the participants of the experiment through written questionnaires. These training data that we can get from experience are given as feedback to the mapping, which is originally trained only with data extracted from sociological studies. This model is generic: it is potentially implementable on any robot, with the exception of the robot-specific MMM mapper.

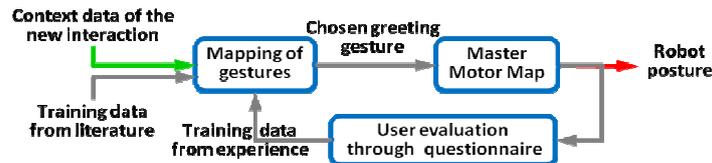


Figure 2. Overview of the greeting selection system. Green arrow: input; red arrow: output.

2.3 Mapping algorithm

Mappings can be trained with data taken from literature of sociology studies. We used data from [5], [16, 17] among others. Training data should be classified through some machine learning method or formula; nevertheless, data taken from these studies feature some properties that may limit the possible choice of classifying methods. In particular, their incompleteness: the focus of sociology papers is set only on specific aspects (such as gender-related studies, which do not provide any information regarding Power relationship) and the resulting data, put into a table, has some missing parts. Missing data makes inconvenient to use techniques, among others, like Principal Component Analysis or Neural Networks. Another constraint in the choice of the method is that context variables are categorical values: they cannot be assigned values like 0, 1, and 2, because applying a mapping method that assumes that $0 < 1 < 2$ would falsify the results.

Considering all these limitations, we decided to use conditional probabilities: in particular Naive Bayes formula, to map the data. The Naive Bayes classifier applies the 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 1 for the class variable C_j and the features x_k from the set X . Our modified version of the classifier takes also into account the possibility of missing data, assigning less weight to them.

$$p(C_j | X) \propto p(C_j) \prod_k p(x_k | C_j) \quad (1)$$

The algorithm also includes rewards or penalties depending on the feedback collected from the experience (namely, participants' questionnaires). This was done because 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. The whole concept of the algorithm is shown the following pseudo-code:

```

begin
D ← training_data; //the dataset (a table containing weights w) is built
for (each participant) begin
  f* ← new_input_context_data; //a vector containing the current context
  if (∃ f* in D) then begin //is f* already contained in the dataset?
    P_f*_g* ← w_f*; //classification directly through weights w for f*
    g* ← argmax(w_f*); //g* is the greeting with the maximum weight
  end;
  else begin
    P_f*_g* ← Naive_Bayes(f*); //probability calculated through Naive Bayes
    g* ← argmax(P_f*_g*); //g* is the greeting with the maximum likelihood
  end;
  bContinueUpdating ← calculate_stopping_conditions();
  if (bContinueUpdating == True) then begin
    eval_g* ← questionnaire_data; //from the participant, on a scale from 1 to 5
    P_f*_g* ← P_f*_g* * (1 + r*1); //positive/negative reward * learning factor
    if (eval_g* ≤ 3) then begin //if evaluation was negative
      g** ← suggested_data //g** is the suggested greeting type appropriate for f*
      P_f*_g** ← P_f*_g** * (1 + 1); //its vector gets a positive reward
      f** ← suggested_data //f** is the suggested context where g* is appropriate
      P_f**_g* ← P_f**_g* * (1 + 1); //its vector gets a positive reward
    end;
    update_dataset(D);
  end;
end; end;

```

The stopping conditions consist in calculating: - whether all possible values of all features have been explored; - whether the moving average of the latest 10 state transitions has decreased below a certain threshold, arbitrarily defined as 2 divided by the number of total states. This means that if mapping has already stabilised, no additional learning algorithm will be performed.

3 Implementation on ARMAR-IIIb

The implementation of the set of gestures 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 [18] was used as intermediate passage.

The MMM is a reference 3D kinematic model that provides a unified representation of various human motion capture systems, action recognition systems, and so on. This representation can be subsequently converted to other representations, such as action recognisers, 3D visualisation or implementation on different robots. In this framework, the MMM is the interface for the transfer of motion knowledge between different embodiments.

The kinematic model of MMM is expanded with statistic/anthropomorphic data, such as: segment properties (e.g. length, mass and so on) defined as a function of global parameters (e.g. body height, weight). The body model of MMM is based on Winter's biomechanical model [18]. It contains some joints, such as the clavícula, which are usually not implemented in robots. A conversion module is necessary to perform a transformation between this kinematic model and ARMAR-IIIb kinematic model. The converter used [18] is based on non-linear optimization to maximise the similarity between the demonstrated human movement and the imitation by the robot.

The simplest and ideal way to reproduce a movement from given joint angles would consist in a one-to-one mapping. However, due to the differences in the kinematic structures of a human and the robot, one-to-one mapping can hardly show acceptable results in terms of humanlike appearance of the movement. In this converter, this problem is addressed by applying a post-processing procedure in joint angle space. 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. A feasible solution is estimated by using the joint configuration of the model on the robot, which serves as an initial solution for a further optimisation step.

We programmed the postures directly on the MMM model (Figure 3, left), and processed them by the converter. As the human model contains many joints, like pelvis, and clavícula, which are not present in the robot configuration, the conversion was not trivial.

The results we obtained with this algorithm needed some retouch, due to some part of the body (e.g. the neck) not implemented in the algorithm.

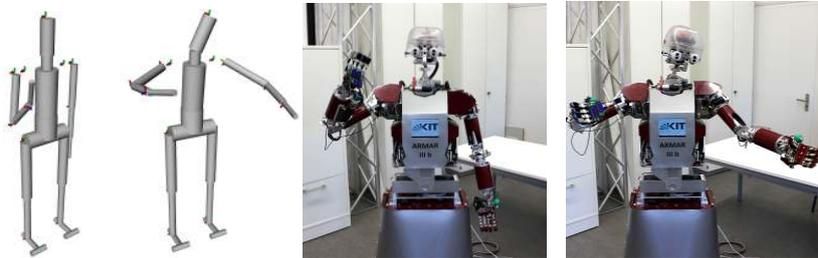


Figure 3. MMM model (left) and implementation on ARMAR (right) of Raise hand and Hug

4 Description of the experiment

4.1 Participants

The experiment was performed in Germany. Participants were 18 German people of different age, gender, workplace, knowledge of the robot, 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 participants repeated the experiment more than once. In this way, we could collect more data, just manipulating the value of one feature: e.g. for the Social distance feature: a participant who meets the robot for the first time can repeat the experiment later on, and will be considered "Acquaintance" instead of "Unknown".

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

4.2 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 sociology data and a mapping M0J was built. After interacting with German people, the resulting mapping M1 was expected to synthesise the rules of greeting interaction in Germany. A mapping M0G made from German sociology data was built but used only for verification.

The experiment protocol was as follows:

- Step 1: The mapping 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. in a public space, meeting for the first time, etc.).
- Step 3: The participant enters the room shown in Figure 4. A curtain covers the location of the robot in order to avoid one of the two parties initiating greeting from a distant location.
- Step 4: Turning left, the participant will find him/herself face to face with the robot, about 2 meters distant. The robot greeting is triggered by an operator as the human participant approaches. The possible choices are: [Bow / Nod / Raise hand / Handshake / Hug]. The two parties have greeted each other
- Step 5: The robot is turned off, and the participant fills questionnaire made of differential semantic scales assessing whether the chosen greeting was appropriate in the actual context. Further details are provided in the algorithm in Section 2.3.

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

Steps 2-6 to be repeated for each participant until stopping conditions are satisfied.

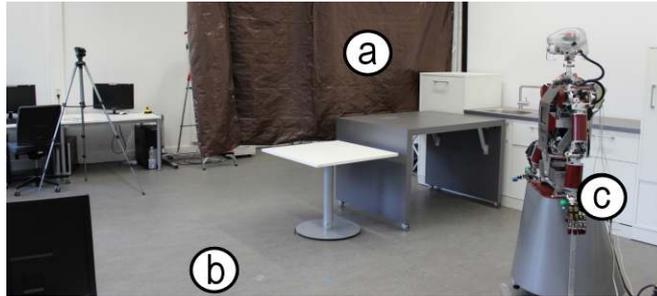


Figure 4. Setup of the room of the experiment. The participant turns left after the curtain that covers the entrance (a) and finds him/herself in (b) face to face with the robot (c).

4.3 Results

The experiment was carried out through 30 interactions, when the moving average of state changes decreased below the threshold, and all greeting gestures 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.

In Tables 1 and 2 it is possible to see the evolution of the mapping of gestures. The counter T , defined as the current number of learning iterations, corresponds to the steps 2 to 6 of the experimental protocol.

TABLE 1. M0J: MAPPING FOR $T=0$

		public male	public female	workp. male	workp. female
close	inf.				
close	equal	green/horiz	green/horiz		
close	sup.	yellow/vert	yellow/vert	yellow/vert	yellow/vert
acquain.	inf.				
acquain.	equal	green/horiz	green/horiz	blue/diag	yellow/vert
acquain.	sup.	yellow/vert	yellow/vert	yellow/vert	yellow/vert
unknown	inf.				
unknown	equal	green/horiz	green/horiz	yellow/vert	yellow/vert
unknown	sup.	yellow/vert	yellow/vert	yellow/vert	yellow/vert

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

TABLE 2. M1: MAPPING FOR $T=30$

		public male	public female	workp. male	workp. female
close	inf.	green/horiz			
close	equal	green/horiz	red/grid	green/horiz	
close	sup.	green/horiz	red/grid		yellow/vert
acquain.	inf.	green/horiz			
acquain.	equal	blue/diag	blue/diag		blue/diag
acquain.	sup.	blue/diag	blue/diag	blue/diag	yellow/vert
unknown	inf.	blue/diag	blue/diag		
unknown	equal	blue/diag	blue/diag	blue/diag	blue/diag
unknown	sup.	blue/diag	blue/diag	blue/diag	yellow/vert

Yellow/vertical lines: bow; grey: nod; blue/diagonal lines: handshake; green/horizontal lines: raise hand; red/grid: hug

This new mapping was verified through an objective function V described in equation 2, which compares two different mappings $M1$ and $M2$.

$$V = \sum_f \sum_j \left(w_j^{(M1_f)} - w_j^{(M2_f)} \right)^2 \quad (2)$$

The function calculates the sum of the variance between the weights w in the same cell f (namely, every possible input value) in two different mappings M1 and M2. Each variance in the weights is calculated not only comparing the greeting with maximum likelihood, but considering the sum of the variances for each greeting j .

The function applied to M0J (Japanese initial mapping) and M1 (final mapping) gives 0.636 as result. Instead, comparing M1 with M0G (German initial mapping) we get 0.324. The t-test of the variances for each f proves the difference to be significant ($p < .05$). This result supports the evolution of mapping M1 from M0J towards M0G.

4.4 Discussion

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. Hug, not present in the Japanese mapping, appears after some participant expressed their feedback indicating that hugging would be appropriate.

Another observation is related to patterns present in the mappings: judging from the patterns in the rows in Table 1, 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 in Japan. 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. After the end of learning phase, the robot can now potentially use two different mappings with human partners of different nationality.

5 Conclusion

In human-robot interaction, it is important for a robot to greet using gestures that are appropriate to specific human cultures in order to improve acceptance and reduce discomfort. For this reason, a system for greeting selection was made. From sociology studies, relevant context features were selected and an algorithm was created to update the mapping that selects the best gesture for each context. Gestures were implemented on the humanoid robot ARMAR-IIIb through the Master Motor Map framework 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 and introduces a model of greetings that can be used with other robots. Ideally, robots will be able in the future to switch between different modes depending on the human cultural background. Future work can carry on towards different directions. Implementation on other robot platforms, and even non-human-like embodiments could be considered. Humanoid robots could

be varying in shape, size and capabilities: using lights, playing sounds and so on. Different channels of communication could lead to different strategies of greeting.

Acknowledgements. This study was conducted as part of the Research Institute for Science and Engineering, Waseda University, and as part of the humanoid project at the Humanoid Robotics Institute, Waseda University. The experiment was carried out in Karlsruhe Institute of Technology, thanks to InterACT, the Waseda/KIT exchange network. We thank all staff and students involved for the support received.

References

1. Nisbett, R.E.: The geography of thought: how Asians and Westerners think differently--and why. Free Press, New York (2004).
2. Rogers, E.M.: Diffusion of Innovations, 5th Edition. Free Press (2003).
3. Trovato, G., Zecca, M., Sessa, S., Jamone, L., Ham, J., Hashimoto, K., Takanishi, A.: Cross-cultural study on human-robot greeting interaction: acceptance and discomfort by Egyptians and Japanese. *Paladyn. Int. Journal of Behavioral Robotics*. 4, 83–93 (2013).
4. Hoffman-Hicks, S.: The longitudinal development of French foreign language pragmatic competence. Doctoral Dissertation, Indiana University (1999).
5. Friedman, H.S., Riggio, R.E., Di Matteo, M.R.: A classification of nonverbal greetings for use in studying face-to-face interaction. *JSAS Catalog of Selected Documents in Psychology*. 11, 31–32 (1981).
6. Hickey, L., Stewart, M.: Politeness in Europe. *Multilingual Matters* (2005).
7. Scherer, S.E., Schiff, M.R.: Perceived intimacy, physical distance and eye contact. *Percept Mot Skills*. 36, 835–841 (1973).
8. Riggio, R.E., Friedman, H.S., DiMatteo, M.R.: Nonverbal Greetings: Effects of the Situation and Personality. *Personality and Social Psychology Bulletin*. 7, 682–689 (1981).
9. Marshall, T.C.: Cultural differences in intimacy: The influence of gender-role ideology and individualism-collectivism. *J of Social and Personal Relat*. 25, 143–168 (2008).
10. Brown, P., Levinson, S.C.: Politeness: Some Universals in Language Usage. Cambridge University Press (1987).
11. Asfour, T., Regenstein, K., Azad, P., et al.: ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control. *Humanoids 2006*. pp. 169–175 (2006).
12. Sakagami, Y., Watanabe, R., Aoyama, et al.: The intelligent ASIMO: system overview and integration. *IROS 2002* pp. 2478–2483 vol.3 (2002).
13. Bum-Jae, Y.: Network-Based Humanoids ‘MAHRU’ As Ubiquitous Robotic Companion. Presented at the 17th IFAC World Congress (2008).
14. Yamamoto, M., Watanabe, T.: Time delay effects of utterance to communicative actions on greeting interaction by using a voice-driven embodied interaction system. *Int. Symp. on Computational Intelligence in Robotics and Automation, 2003*. pp. 217–222 vol.1 (2003).
15. Suzuki, S., Fujimoto, Y., Yamaguchi, T.: Can differences of nationalities be induced and measured by robot gesture communication? *HSI 2011*. pp. 357–362 (2011).
16. Greenbaum, D.P.E., Rosenfeld, H.M.: Varieties of touching in greetings: Sequential structure and sex-related differences. *J Nonverbal Behav*. 5, 13–25 (1980).
17. Sugito, S.: Aisatsu no kotoba to miburi (あいさつの言葉と身振り). *Bunkachou* (1981).
18. Do, M., Azad, P., Asfour, T., Dillmann, R.: Imitation of human motion on a humanoid robot using non-linear optimization. *Humanoids 2008*. pp. 545–552 (2008).