A Novel Greeting Selection System for a Culture-Adaptive Humanoid Robot

Regular Paper

Gabriele Trovato[1], Massimiliano Zecca[2], Martin Do[3], Ömer Terlemez[3], Masuko Kuramochi[4], Alexander Waibel[5,6], Tamim Asfour[3] and Atsuo Takanishi[7*]

1 Faculty of Science and Engineering, Waseda University, Japan
2 School of Electronic, Electrical and Systems Engineering, Loughborough University, UK
3 Institute of Anthropomatics, Karlsruhe Institute of Technology, Germany
4 Kanda Institute of Foreign Languages, Tokyo, Japan
5 Interactive Systems Lab, Karlsruhe Institute of Technology, Germany
6 Language Technologies Institute, Carnegie Mellon University, Pittsburgh, USA
7 Department of Modern Mechanical Engineering, Waseda University; Humanoid Robotics Institute (HRI), Waseda University, Japan
* Corresponding author(s) E-mail: contact@takanishi.mech.waseda.ac.jp

Abstract

Robots, especially humanoids, are expected to perform human-like actions and adapt to our ways of communication in order to facilitate their acceptance in human society. Among humans, rules of communication change depending on background culture: greetings are a part of communication in which cultural differences are strong. Robots should adapt to these specific differences in order to communicate effectively, being able to select the appropriate manner of greeting for different cultures depending on the social context. In this paper, we present the modelling of social factors that influence greeting choice, and the resulting novel culture-dependent greeting gesture and words selection system. An experiment with German participants was run using the humanoid robot ARMAR-IIIb. Thanks to this system, the robot, after interacting with Germans, can perform greeting gestures appropriate to German culture in addition to a repertoire of greetings appropriate to Japanese culture.

1. Introduction

Acceptance of humanoid robots in human societies is a critical issue. One of the main factors is the relationship between the background culture of human partners and acceptance. According to the traditional view in literature, anxiety towards robots is more common in Western countries. In fact, differences between East and West in cognition — due to differing ecologies, social structures, philosophies, and educational systems — can be traced back to the ancient cultures of China and Greece [1]. However, this traditional view has been debated, as there are positive examples of robotic heroes in Western science fiction; technology acceptance, for instance, also depends on the country that is the producer, since the culture of that country may create bias towards some aspects of the

product. As a consequence, localization of products may occur. It is then necessary to understand cultural norms of the country for ensuring technology acceptance [2,3].

In the work of Trovato et al. [4], culture-dependent acceptance and discomfort relating to greeting gestures were found in a comparative study with Egyptian and Japanese participants. As the importance of culture-specific customization of greeting was confirmed, the need for a system of greeting selection for robots was highlighted. In other words, acceptance of robots can be improved if they are able to adapt to different kinds of greeting rules.

Adaptive behaviour in robotics can be achieved through various methods: most commonly by reinforcement learning (such as in [5]), but also via neural networks [6], genetic algorithms [7], and function regression [8] among others. Various approaches are possible, but the understanding and reproducing of adapting human-like or animal-like behaviours remains a challenge [9].

1.1 Greeting interaction with robots: related works

Robots, especially humanoids and anthropomorphic ones, are expected to interact and communicate with humans of different cultural background in a natural way. It is therefore important to study greeting interaction between robots and humans.

Some humanoid robots can perform programmed greetings: among others, ARMAR-III [10], which greeted the Chancellor of Germany, Angela Merkel, with a handshake. ASIMO [11] is capable of performing a wider range of greetings: a handshake, waving both hands, and bowing, and can recognize such gestures performed by others. HRP-4 [12] and MAHRU [13] are two other examples of humanoid robots that can greet by a simple bow.

While greeting gestures have been implemented, so far only a few greeting interaction experiments with robots have been conducted in order to test the impression on humans. Examples are experiments by Yamamoto et al. [14], who focused on timing, rather than on culture, and experiments featuring the social robot ApriPoco, in which Japanese, Chinese, and French greetings were compared [15]. However, in experiments with ApriPoco, conclusions remain unclear because of the small number of subjects and the limited number of degrees of freedom of the robot, leading to difficulties in obtaining significant data from biological signals from the test subjects. Compared to those experiments, our intention is to do a more extensive study, with a greater number of subjects and a human-sized anthropomorphic robot.

1.2 Objectives of this paper

The robot should be trained with sociology data related to one country, and evolve its behaviour by 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, designed for close cooperation in human

environments. As the experiment is carried out in Germany, the interactions are with German participants, while preliminary training is done with Japanese data, which is culturally extremely different.

The idea behind this study is a typical scenario in which a foreigner visiting a country 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 this experiment, we want a robot to be able to do the same.

This work is an application of a study of sociology into robotics. Our contribution is to synthesize the complex and sparse data related to greeting types into a model; create a selection and adaptation system; and implement the greetings in a way that can potentially be applied to any robot.

Other existing studies focus on specific subfields of greetings: sociology studies focus on specific greetings or on the effect of specific factors; whereas robotics studies like [16,17] focus more on physical aspects, such as the oscillation trajectory of a handshake. Our approach was related to the psychological aspect of greeting interaction, and the scope of the study was more extensive, focusing less on the details of each single greeting.

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; in Section 5 we discuss the results; and in Section 6 we conclude the paper.

2. Greeting selection

2.1 Greetings among humans

Greetings are the means of initiating and closing an interaction. Hoffman-Hicks [18] states that greetings function primarily as formulaic exchanges, which serve to acknowledge another person's presence. We desire that robots be able to greet people in a similar way to humans. For this reason, understanding current research on greetings in sociological studies is necessary. Moreover, depending on cultural background, there can be different rules of engagement in human-human interaction. Gaps in recognition of facial expressions and gestures due to a lack of understanding regarding cultural norms can lead to difficulty in communication. For example, the complexity of greetings in Japanese culture may cause possible communication problems with foreigners [19].

A unified model of greetings does not seem to exist in the literature, but a few studies have attempted a classification of greetings. Some more specific studies have been done on handshaking [20]. Varieties of bowing also exist: this is why

many publications advising foreigners doing business in Japan exist [21].

A classification of greetings was first attempted by Friedman [22] based on intimacy and commonness. The following greeting types were mentioned: smile; wave; nod; kiss on mouth; kiss on cheek; hug; handshake; pat on back; rising; bow; salute; and kiss on hand. Greenbaum et al. [23] also performed a gender-related investigation, while [24] contained a comparative study between Germans and Japanese.

Many other contributions do not attempt to list greeting types or to classify them, but to 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.

As Spencer-Oatey pointed out [25], authors often use the same terms with different meanings, or different terms with the same meaning. We will try to keep the terms consistent. For instance, context is a word that is sometimes used, but its actual meaning denotes location (private or public). As Sugito [24] and Firth [26] mentioned, location influences intimacy and greeting words; we 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 the degree of contact. In fact, intimacy can be intended to denote closeness in contact, or a closeness of acquaintanceship: we will use the term 'intimacy' to represent the closeness in contact, while closeness of acquaintanceship will be described by 'social distance'. According to the literature, intimacy is influenced by physical distance, eye contact [27], gender [28,29], location [22], and culture [30].

Politeness [31,32] is a key concept in sociology: Brown and Levinson were the first to create a formula for calculating it. Even though they did not numerically define a coefficient, they represented politeness as a function of power balance in a relationship, social distance, and as a cultural factor [31]. Affect (by Slugoski [33]) can be included in this formula, too, but it is usually comprehended within social distance. Other factors that influence politeness were defined by Ferguson [34]: the number of individuals and the time passed since the previous interaction. Kern and Eichmueller [35] mention the same factors and some others, including age, which directly influence the choice of greeting words. Some more factors described by Li [36] influence only greeting words: time, regionality, setting, and content.

'Time' is a factor distinguishing the use of greetings, e.g. in case of seasonal greetings, introductory greetings, and ceremonial greetings. In particular, time of the day is important for the choice of words [26,36,37]. 'Setting' indicates greetings conducted though telephones, TVs, 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 current terms used in sociology (Figure 1) will be used to develop a comprehensive greeting selection model.

2.2 Model of greetings

It is clear that as it is, the graph we made is too complex to be usable. It needs to go through a process of simplification: in Figure 1, the factors to be cut are greyed out.
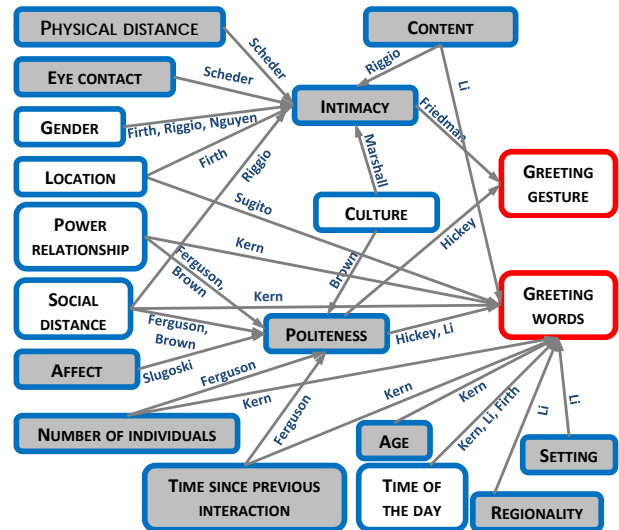

Figure 1. Overview of factors that influence greeting choice. The names on the arrows indicate the authors of relevant publications. Factors coloured in grey are omitted from our study.

The simplification was guided by the following assumptions:

1. Only two individuals (a robot and a human participant): we do not take in consideration a higher number of individuals.

2. Eye contact is taken for granted: as the establishment of eye contact is a problem for machine vision, let us suppose that the two parties meet face to face (and plan the experimental setup accordingly).

3. Age is considered part of 'power relationship': even though they are two distinct factors, putting them together allows us to manipulate the power relationship in an experiment by the inclusion of volunteers of different ages.

4. Regionality is not considered: we will consider standard languages, without taking dialects into account.

5. Setting is not considered: this factor involves the use of devices such as a phone, while the experiment will be face-to-face and no other devices will be used.

6. Physical distance is close enough to allow interaction: as a large or small distance can limit the range of possible gestures, we suppose that in the experiment the two parties will find themselves face-to-face

Gabriele Trovato, Massimiliano Zecca, Martin Do, Ömer Terlemez, Masuko Kuramochi, Alexander Waibel, Tamim Asfour and Atsuo Takanishi: A Novel Greeting Selection System for a Culture-Adaptive Humanoid Robot

3

without obstacles between them, and will leave this factor out.

7. Gender is intended to be a same-sex dyad: in sociology studies, interactions can be divided between same-sex or opposite-sex dyads. As the gender of 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.

8. Affect is considered together with 'social distance': this is the standard interpretation, as in Brown and Levinson's Politeness [31].

9. Time since the last interaction is partially included in 'social distance': meeting for the first time, rather than meeting after a short or long time, certainly makes a difference in the manner of greeting. However, if we simplify the time measurement 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 an (undefined) time".

10. Intimacy and politeness are not necessary: they are two key concepts in sociology, but they are intermediate passages from the upstream factors and the downstream result. For this reason, both factors can be eliminated as they are considered implicit in all these correlations.

2.3 Greeting selection system

After the simplifications described in the previous paragraph, the resulting factors could be summarized in Figure 2. 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 two or three values.

The outputs can also assume only a limited set of categorical values, the classes of a mapping problem. The greeting gestures list has been defined from the relevant sources mentioned above [22–24]. Originally, the set contained six gesture types, including 'kiss', which was dropped because it was 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. The greeting words list has been defined by selecting the most common greeting words and getting information from relevant studies [24,37].

Figure 3 contains the overview of the greeting model. It takes context data as input and produces the appropriate robot posture and speech for that input. The context is the set of features shown in Figure 2. The mapping is updated by an algorithm that will be described in Section 2.5. Two
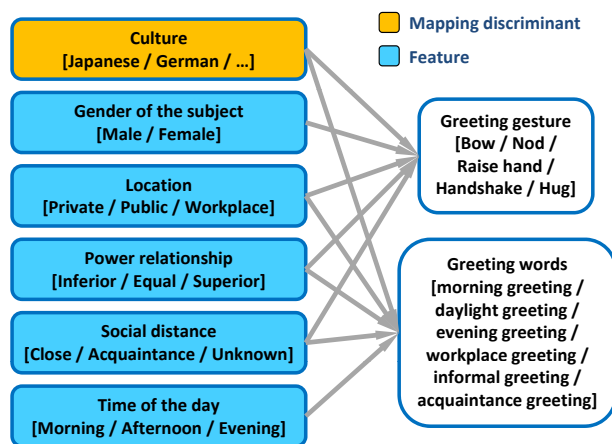

Figure 2. Features, mapping discriminants, classes, and their possible states

different mappings are made, one for gestures and one for words. Both mappings give as an output the most appropriate selection.

As shown in the right-hand side of the graph in Figure 3, words are turned into speech through a freeware text-to-speech software, and the speech file is then played through the speakers of the robot. The chosen gesture is turned into a robot configuration through the Master Motor Map [38], which will be described in more detail in Section 3.
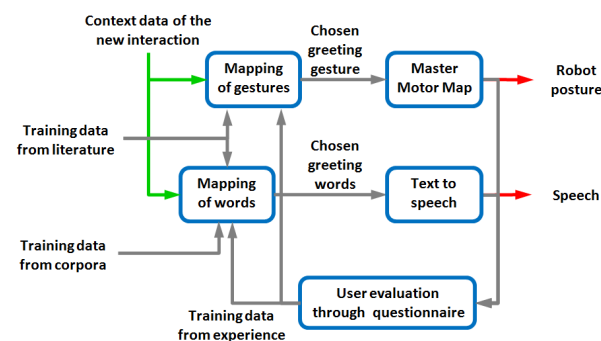

Figure 3. Overview of the greeting selection model

The two outputs are evaluated by the participants of the experiment through written questionnaires. These training data that we get from the experience are 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 the Master Motor Map, which is a component that could be omitted if robot gestures are programmed manually.

2.4 Greeting selection system training data

Mappings can be trained to an initial state with data taken from the literature of sociology studies. Training data should be classified through some machine learning method or

formula; nevertheless, data taken from these studies have some properties that limit the possible choice of classifying methods. An issue with these studies is 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 relationships, for example. The resulting data, put into a table, has some missing parts. This is true for both gestures and words.

Considering these limitations, we decided to use conditional probabilities: in particular the 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 for 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 also takes into account the possibility of missing data, assigning less weight to them.

$$p\left(C_j \mid X\right) \propto p\left(C_j\right) \prod_k p\left(x_k \mid C_j\right) \qquad (1)$$

While training data of gestures can be obtained from the literature, data of words can also be obtained from text corpora. In linguistics, a corpus is a large and structured set of texts. Sometimes, portions of speech recordings are transcribed into a corpus and then analysed. Corpora are then used to perform 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 [39], where the Suprasegmental Hidden Markov Model is applied to detect emotions in speech.

English corpora, such as the British National Corpus, or the Corpus of Historical American English, are readily available online. Using such online tools, it is possible to analyse greeting words usage depending on the context. For example, counting the occurrence of a greeting word ("hello" or "good morning") together with some indication of a distant relationship (Mr., Dr.) or a close relationship ("darling", etc.). In a similar way, analysis of the other features ('time of the day', 'gender', etc.) can be carried out.

In Table 1, 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 the 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 terms of the 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 (column AB); their span; and the last column contains an index called Mutual Information (MI). 'Span' is the range between the two words in order to be accounted as in close proximity to one another : '1' indicates

that the word pair is counted whenever the words occur next to one another; '2' indicates that there is one word in between, and so on. Span is also a parameter for the calculation of the index *MI*, which is defined as in equation 2, as adaptation to the corpora of the generic information theoretic measure [38]. *MI* assesses the correlation between two words: the higher the value, the higher the strict correlation between the two words. Words to analyse in association with greetings are useful when the context of the corpora cannot be determined a priori precisely.

Words A B	sizeCorpus	A	B	AB	span	MI
Hello darling	8076643	2287	600	27	1	7.31
Hello love	8076643	2287	2553	17	1	4.56

Table 1. Example of data extracted from corpora

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

While English corpora are relatively easy to analyse, Japanese ones are trickier because, due to the lack of spaces between words in the Japanese language, it is often impossible for analysis tools to determine where a word ends and the next word begins. This fact makes calculation of span, and therefore of the *MI*, tricky or inaccurate. As a manual revision of huge amounts of text would be necessary for solving this problem, leading to a drift away from the scope of this research, we decided to not rely on corpora for Japanese language. Conversely, training data of Japanese greeting words were extracted from relevant Japanese sociology studies [24,37,41–43]. Japanese training data for gestures were extracted from [24,44–46], while the training sets for gestures in other Western countries was made using data from [22,23,47].

In sociology papers, raw data are often reported in the form of percentages. These numbers can be easily converted into weights between zero and one. For words, weights can be obtained from their rate of occurrences.

Features are determined from the characteristics of the sociological study (for example, if the study is on the use of handshakes among men and women, the feature 'gender' will be set) or of the context of the dialogue in the corpus (for instance, in a dialogue between two friends, 'social distance' will be set as 'close'). Non-specified information will be left as blank in the training data, and the algorithm will take missing data into account. For example, this would happen in case of extracting data from a study focused on greetings and power relationships, which does not take gender into account.

In the present study, the location of the experiment was Germany. For this reason, the only dataset needed was the Japanese. As stated in the motivations at the beginning of this paper, the robot should initially behave like a foreigner:

Gabriele Trovato, Massimiliano Zecca, Martin Do, Ömer Terlemez, Masuko Kuramochi, Alexander Waibel, Tamim Asfour and Atsuo Takanishi: A Novel Greeting Selection System for a Culture-Adaptive Humanoid Robot

5

ARMAR-IIIb, trained with Japanese data, will have to interact with German people and adapt to their customs.

2.5 Mapping and questionnaires

The mapping is represented by a dataset, initially built from training data, as a table containing weights for each context vector corresponding to each greeting type. We now need to update these weights.

Whenever a new feature vector (representing context) is given as an input, it is checked to see 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 the Naive Bayes classifier. The output is the chosen greeting, after which the interaction will be evaluated through a questionnaire consisting in the following three questions, to be answered in a five-point semantic differential scale:

1. How appropriate was the greeting chosen by the robot for the current context?

2. (If the evaluation at point 1 was <= 3) which greeting type would have been appropriate instead?

3. (If the evaluation at point 1 was <= 3) which context would have been appropriate, if any, for the greeting type of point 1?

Weights of the affected features are multiplied by a positive or negative reward (inspired by reinforcement learning) which is calculated proportionally to the evaluation (will be negative for a rating that equals 1 or 2). A decreasing learning factor also affects the reward.

Mappings stop evolving when the following two stopping conditions are satisfied: all possible values of all features have been explored; and the moving average of the latest 10 state transitions has decreased below a certain threshold.

Thanks to this implementation, mappings can evolve quickly, without requiring hundreds or thousands of iterations, but rather a number comparable to the low number of interactions humans need to understand and adapt to social rules.

3. Implementation on ARMAR-IIIb

ARMAR-III [12] is a 43 degrees of freedom robot that is designed for close cooperation with humans. Therefore, the robot has a humanlike appearance and its purpose is to have sensory capabilities similar to humans. ARMAR-IIIb is a slightly modified version with different shape to the head, the trunk, and the hands.

3.1 Implementation of gestures

The implementation on the robot of the set of gestures defined in Figure 2 was performed in a way that it is not strictly hardwired to the specific hardware. Rather than manually defining the patterns of the gestures, the Master Motor Map (MMM) was used as an intermediate passage.

The 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 recognizers, 3D visualization, or implementation into different robots. In the framework proposed in [31] and shown in Figure 4, the MMM is the interface for the transfer of motion knowledge between different embodiments. The MMM is intended to become a common standard in the robotics community, to allow the establishment of common benchmarks and the sharing of 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 [38]. It is made by setting a maximum number of DoF that might be used by any visualization, recognition, or reproduction module.

The body model of MMM based on Winter's biomechanical model can be seen in the left-hand illustration in Figure 5. It contains some joints, such as the clavicula, which are usually not implemented in humanoid robots. A conversion module is necessary to perform a transformation between this kinematic model and ARMAR-IIIb kinematic model, in the right-hand illustration in Figure 5.

The converter we used [48] is a module that was created for making imitation learning tasks easier. It is based on non-linear optimization to maximize 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 between an observed human subject and the robot. However, due to the differences in the kinematic structures of a human and the robot (such as measurements of joints and limb), one-to-one mapping can hardly show acceptable results in terms of a human-like 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 optimized 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 MMM model on the robot, which serves as an initial solution for a further optimization step. This ensures human-like motion in the robot.

The MMM framework is mainly used for humanoid robots and has a high support for every kind of human-like robot. Using human-like robots, transfer rules can be defined,
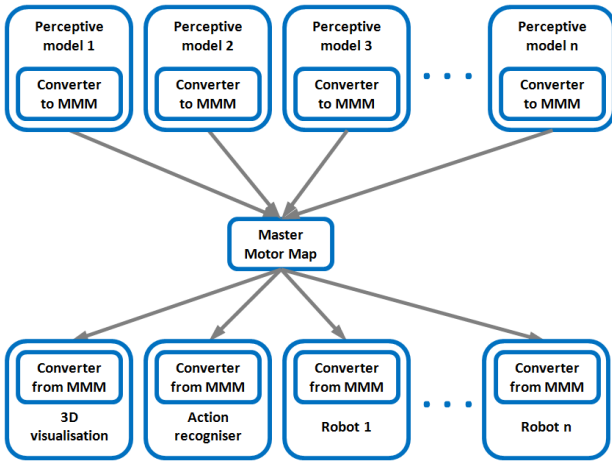
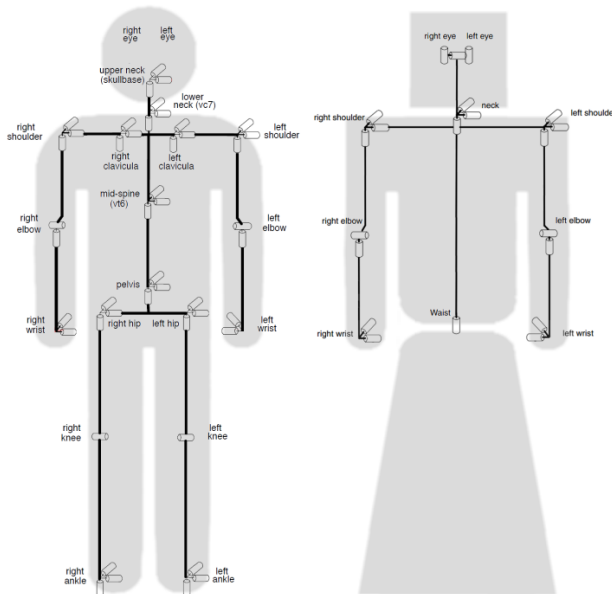**Figure 4.** Illustration of the Master Motor Map framework

Figure 5. Body model of Master Motor Map and ARMAR-IIIb configuration

which can convert human motion from the MMM reference model to the robot. This can be either the whole body motion of the human or only parts of it, for example only the upper body or only the right arm. A non-human-like robot model with an arbitrary number of DoFs is also supported, however it may not be possible to find a good conversion from the MMM reference model to that specific robot. In that case, however, the motion representation parts of the framework can be used nevertheless.

After programming the postures directly on the MMM model (Figure 6), they were processed by the converter. As mentioned previously, the human model contains many joints, such as the 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 refined, due to some part of the body (e.g., the neck) not being implemented in the algorithm. In Figure 7, the final result is shown.

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 8, 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.
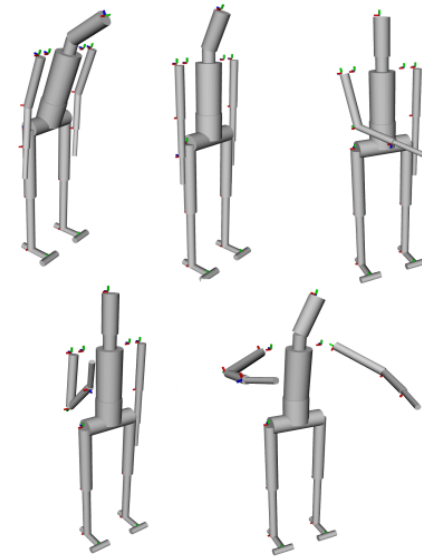

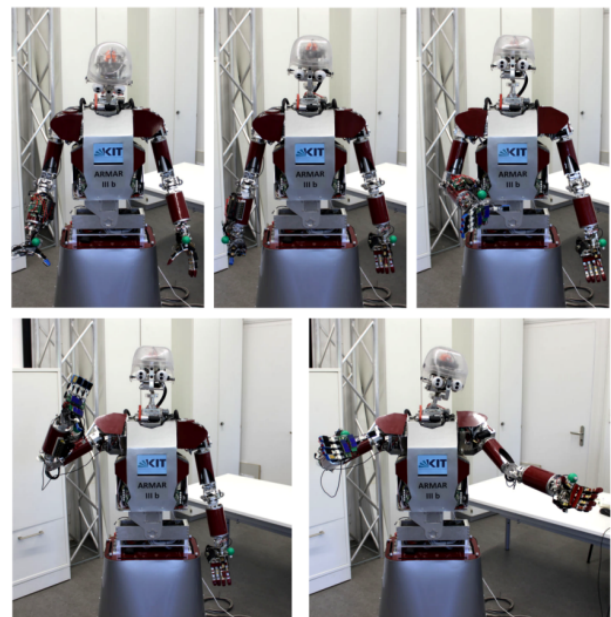**Figure 6.** Output gestures: MMM model. Top row: bow, nod, handshake. Bottom row: raise hand, hug.

Figure 7. Output gestures: implementation on ARMAR-IIIb. Top row: bow, nod, handshake. Bottom row: raise hand, hug.
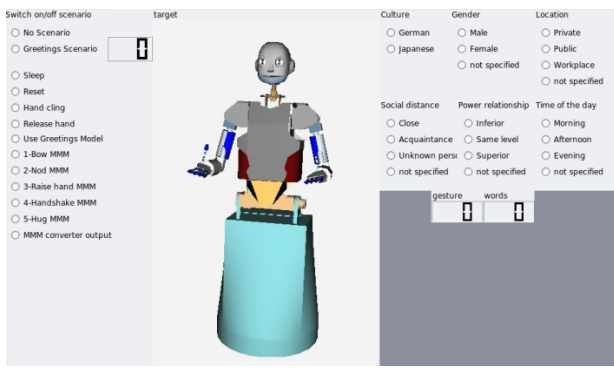
Gabriele Trovato, Massimiliano Zecca, Martin Do, Ömer Terlemez, Masuko Kuramochi, Alexander Waibel, Tamim Asfour and Atsuo Takanishi: A Novel Greeting Selection System for a Culture-Adaptive Humanoid Robot

7

Figure 8. MCA Interface for the control of ARMAR-IIIb

3.2 Implementation of words

As seen in Figure 2, the possible options of output words have been defined. This set of greetings has been translated into both German and Japanese, as in Table 2, 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 a 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 dictionary terms, but does not understand the difference in usage of these words in different contexts.

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

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

Table 2. Conversion table of greeting words

4. Experiment description

4.1 Participants

The experiment was performed in Germany in the room shown in Figure 9. Participants were 18 German people of different ages, genders, workplaces, and knowledge of the robot, in order to ensure that the mapping could be trained with various combinations of context.

It was not possible to include all combinations of feature values 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 by manipulating the value of a single feature: for instance, 'time of the day', when the experiment is repeated at different times, or 'social distance', which becomes 'acquaintance' rather than 'unknown' during later interactions.

The demographics of the 18 participants were as follows: M: 10; F: eight; average age: 31.33; age standard deviation: 13.16. However, the number of interactions was determined by the stopping condition of the algorithm.
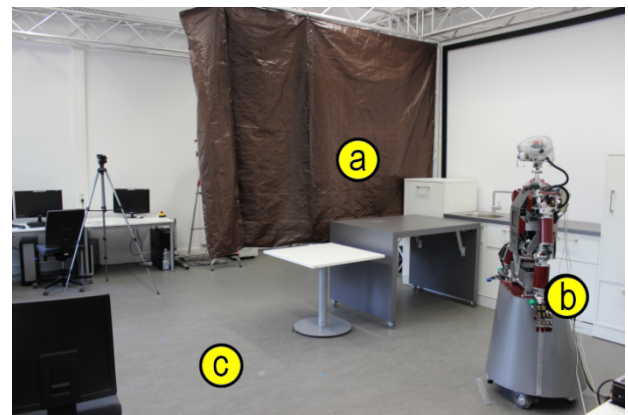

Figure 9. Setup of the room of the interaction experiment

The number of interactions, taking repetitions into account, was 30: M: 18; F: 12; average age: 29.43; age standard deviation: 12.46.

4.2 Experiment 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 two mappings M0J were built for gestures and words, respectively. After interacting with German people, the resulting mappings M1 were expected to synthesize the rules of greeting interaction in Germany. A mapping M0G of gestures, made from German sociology data was built but used only for verification. A mapping of German words could not be made as it would have required a study of German corpora, which is out of the scope of this research.

The experiment protocol is as follows:

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

Step 2. Contextual data about the encounter are given as inputs to the algorithm and the robot is prepared. In the meantime, the participant is instructed to enter the room, turn left, and greet the robot naturally with consideration to the current context (e.g., it is morning, in a public space,

they are meeting for the first time, etc.) without having to match the robot's own greeting type.

Step 3. The participant enters the room shown in Figure 9. A curtain (a) covers the location of the robot (b), therefore the participant will find him/herself in the location (c) face-to-face with the robot, about 2 metres distant. In this way, both the possibility of greeting from a distant location and of asynchronous greetings by the two parties are avoided.

Step 4. The robot's greeting is triggered by an operator as the human participant approaches (Figure 10). Both gesture and words get triggered at the same time.

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.

Step 6. The mapping is updated using the subject's feedback (see Section 2.5). The new mapping will be used in the next interaction.

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

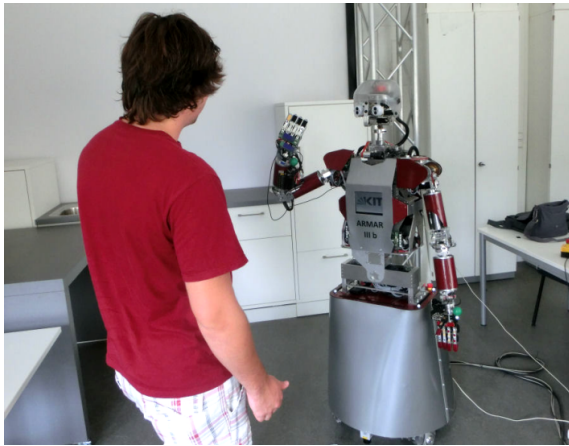**Step 8.** Training stops after the state changes are stabilized.

Figure 10. Viewpoint of the participant face-to-face with ARMAR

4.3 Results

The experiment was carried out through 30 interactions, and all greeting gestures and word types had the chance of being selected at least once. The features 'gender' was explored 34 times; 'location' 50 times; 'power relationship' 56 times; 'social distance' 46 times; 'time of the day' 39 times.

Handshakes, shown in Figure 11, were common after the mapping started to change. In Table 3 it is possible to see the evolution of the maximum likelihood, in the mapping of gestures. The counter T, the current number of learning iterations, corresponds to steps 2 to 6 of the experimental protocol. The algorithm stopped running after 30 iterations, when both the moving averages of state changes decreased below a certain threshold.

The new mapping of gestures was verified through an objective function V described in equation 3, which compares two different mappings M1 and M2.

$$V = \sum_f \sum_j \left(w_j^{(M1_f)} - w_j^{(M2_f)} \right)^2 \qquad (3)$$

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.

It can also be noted from the evolution of mapping, that after the interactions, the states in which bowing is preferred has greatly decreased, while handshake has become more common. On the other hand, the bow has not disappeared. Hugs, not present in the Japanese mapping, appeared after some participants expressed feedback indicating that hugging would be appropriate.

The evolution of maximum likelihood in the mapping of words is shown in Table 4. The main change is the disappearance of the workplace greeting in German mapping. We also expected other subtle changes to be noticeable. Among them were the more common use of an informal greeting by Germans: however, the changes cannot be considered significant. Conversely, some other patterns can be found in the gestures' mappings: judging from the columns in Table 3 for T = 0, it is clearly visible that a strict categorization is present in the Japanese mapping with regards to social distance, whereas the same pattern is not present in the German mapping.

Labels top row: Social distance; second row: Power relationship. Labels left column: Location; second column: Gender.

Bow — Nod — Handshake — Raise hand — Hug

Table 3. Evolution of mapping of gestures: maximum likelihoods for M0J and M1

Gabriele Trovato, Massimiliano Zecca, Martin Do, Ömer Terlemez, Masuko Kuramochi, Alexander Waibel, Tamim Asfour and Atsuo Takanishi:
A Novel Greeting Selection System for a Culture-Adaptive Humanoid Robot

9

T = 0	mor inf	mor eq	mor sup	aft inf	aft eq	aft sup	eve inf	eve eq	eve sup
pub clo									
pub acq									
pub unk									
wor clo									
wor acq									
wor unk									
T = 30	mor inf	mor eq	mor sup	aft inf	aft eq	aft sup	eve inf	eve eq	eve sup
pub clo									
pub acq									
pub unk									
wor clo									
wor acq									
wor unk									

Labels top row: Time of the day; second row: Power relationship.
Labels left column: Location; second column: Social distance

- Morning greeting
- Daylight greeting
- Evening greeting
- Informal greeting
- Workplace gr.
- Acquaintance gr.

Table 4. Evolution of mapping of words: maximum likelihoods for M0J and M1
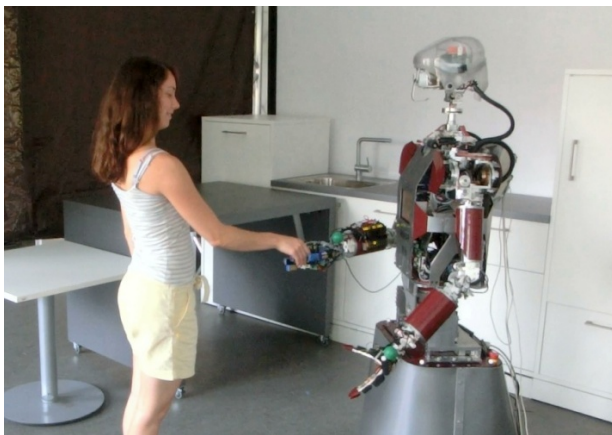

Figure 11. Example of a handshake with ARMAR

This fact seems to accord with the more hierarchical view of society that the Japanese have. Both the resulting German and Japanese mappings may not be 100% accurate compared to reality, but they are a simplification that is consistent with German participants' feedback and Japanese sociology literature, respectively.

5. Discussion

5.1 On this approach

In Section 1.2 we stated our main contributions. Within our work, the concept of a greeting selection system is novel, and while the comprehensive study of current greeting choice factors is also new in sociology, its modelling and application can be useful in robotics.

In particular, one 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.

5.2 Impact on people

The work described in this manuscript is the implementation of cultural adaptation during greeting interaction; the need for such cultural adaptation was highlighted in the work in [4], to which we are making a continuation. The results in that cross-cultural experiment support the hypothesis that people from different cultures prefer a robot, and anthropomorphize it more, if it is culturally close to them, whereas the likeability factor and the number of interactions with a robot drop in the case of a culturally distant robot.

As these findings were confirmed by questionnaires as well as participants' reactions and non-verbal behaviour, we do not show in this paper measurements of the impact on participants, which we will do in a future work. Further information can be found in [4].

5.3 Limitations and improvements

In the current implementation, there are also a few limitations. The first obvious limitation is related to the manual input of context data, with no perception from the robot. The integrated use of cameras would make it possible to determine features such as gender (as in [49]), age, and race of the human. Speech recognition system and cameras could also detect the human's own greeting. Estimating whether the robot's own greeting was appropriate or not, without the help of a human supervisor, can be especially tricky. A cost/reward function that takes into consideration the distance of the other party, the timing of the greeting, head orientation, and other cues could suggest whether the reaction to the greeting matched the expected one. However, this information would still be less accurate than an explicit answer collected through a questionnaire.

Another limitation was the choice of using just literature for training data. 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 that occurred during the definition of the greetings model.

5.4 Different kinds of embodiment

The definition of a set of greeting gestures was based on human-related studies, taking for granted that the humanoid robot has a body that resembles humans and possesses a similar range of motion. However, humanoid robots could vary in shape, size, and capabilities, and this could produce an effect in which greeting types are selected as being more fitting for each robot.

A possible extension of this work could be making a robot discover autonomously the optimal way of attracting the attention and starting an interaction with a human, depending on the characteristics of its own body. Communication channels can in fact rely on visual or auditory aids. For instance, a device such as a mobile phone can initiate

communication with a human through vibration, and a robot that can blink its eyes could use lights and colours to communicate in a dark environment.

6. Conclusion

In this paper, a system of culture-dependent greeting selection was presented. A preliminary survey of up-to-date work in sociology was done in the field of greetings, and from the resulting correlation graph a novel model of greetings was created. It is based on a mapping that can evolve from one culture to another through an updating algorithm, selecting the best gestures and words for each context. Gestures were implemented on the humanoid robot ARMAR-IIIb through the Master Motor Map framework, whereas words were spoken through text-to-speech software. An experiment was performed with German participants: through their feedback, ARMAR-IIIb could successfully learn a new (German) mapping of greetings selection, starting from an initial Japanese mapping. The objective stated at the beginning of the paper was achieved, as the robot, trained using sociology data related to one country, was able to evolve its behaviour in a small number of interactions. This work is a step towards culture-related robot customization; ideally, robots will one day be able to switch easily between different modes depending on the cultural background of the human partner. Future work includes reconsidering the limitation of the implementation of gestures itself, and exploring different strategies of initiating communication depending on the specific embodiment of the robot.

7. Acknowledgements

8. References

[1] Nisbett RE (2004) The geography of thought: how Asians and Westerners think differently — and why. Free Press, New York.

[2] Thomas RM (1987) Computer technology: an example of decision-making in technology transfer. Educational Technology—Its Creation, Development and Cross-Cultural Transfer, R. M. Thomas & V. N. Kobayashi. Oxford: Pergamum Press, pp. 25–34.

[3] Rogers EM (2003) Diffusion of Innovations, 5th Edition, 5th edition. Free Press.

[4] Trovato G, Zecca M, Sessa S, et al. (2013) Cross-cultural study on human-robot greeting interaction: acceptance and discomfort by Egyptians and Japanese. Paladyn International Journal of Behavioral Robotics 4:83–93. doi: 10.2478/pjbr-2013-0006.

[5] Zalama E, Gomez J, Paul M, Peran JR (2002) Adaptive behavior navigation of a mobile robot. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 32:160–169. doi: 10.1109/3468.995537.

[6] Navarro N, Munoz C, Freund W, Arredondo VT (2006) Acquiring Adaptive Behaviors of Mobile Robots Using Genetic Algorithms and Artificial Neural Networks. Electronics, Robotics and Automotive Mechanics Conference, pp. 87–91.

[7] Savage J, Savage R, Morales-Aguirre M, Kuri-Morales A (2012) Adaptive FPGA-based Robotics State Machine Architecture Derived with Genetic Algorithms (Abstract Only). Proceedings of the ACM/SIGDA International Symposium on Field Programmable Gate Arrays. ACM, New York, NY, USA, pp. 267–267.

[8] Banzhaf W, Nordin P, Olmer M (1997) Generating Adaptive Behavior using Function Regression within Genetic Programming and a Real Robot. In 2nd International Conference on Genetic Programming. Morgan Kaufmann, pp. 35–43.

[9] Wynne CDL, Staddon JER (2013) Models of Action: Mechanisms for Adaptive Behavior. Psychology Press.

[10] Asfour T, Regenstein K, Azad P et al. (2006) ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control. 2006 6th IEEE-RAS International Conference on Humanoid Robots. pp. 169–175.

[11] Sakagami Y, Watanabe R, Aoyama C et al. (2002) The intelligent ASIMO: system overview and integration. IEEE/RSJ International Conference on Intelligent Robots and Systems, 3: 2478–2483 vol.3.

[12] Kaneko K, Kanehiro F, Morisawa M et al. (2011) Humanoid robot HRP-4 — Humanoid robotics platform with lightweight and slim body. 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 4400–4407.

[13] Bum-Jae Y (2008) Network-Based Humanoids 'MAHRU' As Ubiquitous Robotic Companion. In: Myung C (ed) pp. 724–729.

[14] Yamamoto M, Watanabe T (2003) Time delay effects of utterance to communicative actions on greeting interaction by using a voice-driven embodied interaction system. 2003 IEEE International Sympo-

sium on Computational Intelligence in Robotics and Automation, 2003. Proceedings. 1: 217–222.

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

[16] Kasuga T, Hashimoto M (2005) Human-Robot Handshaking using Neural Oscillators. Proceedings of the 2005 IEEE International Conference on Robotics and Automation pp. 3802–3807.

[17] Jindai M, Watanabe T, Shibata S, Yamamoto T (2006) Development of a Handshake Robot System for Embodied Interaction with Humans. The 15th IEEE International Symposium on Robot and Human Interactive Communication pp. 710–715.

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

[19] Fan SK (2002) Japanese communication beyond language◎: From the viewpoint of Hong Kong Chinese. The bulletin of the Institute for Japan Studies, Kanda University of International Studies 3:35_a–21_a.

[20] Huwer J (2003) The effect of context, intimacy and gender on handshaking. Senior Thesis Project, Department of Psychology, Haverford College.

[21] Nishiyama K (2000) Doing Business with Japan: Successful Strategies for Intercultural Communication. University of Hawaii Press.

[22] Friedman HS, Riggio RE, Di Matteo MR (1981) A classification of nonverbal greetings for use in studying face-to-face interaction. JSAS Catalog of Selected Documents in Psychology 11: 31–32.

[23] Greenbaum DPE, Rosenfeld HM (1980) Varieties of touching in greetings: Sequential structure and sex-related differences. J Nonverbal Behav 5: 13–25. doi: 10.1007/BF00987051.

[24] Sugito S (1981) Aisatsu no kotoba to miburi. Bunkachou.

[25] Spencer-Oatey H (1996) Reconsidering power and distance. Journal of Pragmatics 26: 1–24. doi: 10.1016/0378-2166(95)00047-X.

[26] Firth R (1972) Verbal and bodily rituals of greeting and parting. In: La Fontaine J (ed) The Interpretation of Ritual: Essays in Honour of A.I. Richards. Tavistock Publications, pp. 1–38.

[27] Scherer SE, Schiff MR (1973) Perceived intimacy, physical distance and eye contact. Percept Mot Skills 36: 835–841.

[28] Nguyen T, Heslin R, Nguyen ML (1975) The Meanings of Touch: Sex Differences. Journal of Communication 25: 92–103. doi: 10.1111/j.1460-2466.1975.tb00610.x.

[29] Riggio RE, Friedman HS, DiMatteo MR (1981) Nonverbal Greetings: Effects of the Situation and Personality. Personality and Social Psychology Bulletin 7: 682–689. doi: 10.1177/014616728174027.

[30] Marshall TC (2008) Cultural differences in intimacy: The influence of gender-role ideology and individualism—collectivism. Journal of Social and Personal Relationships 25: 143–168. doi: 10.1177/0265407507086810.

[31] Brown P, Levinson SC (1987) Politeness: Some Universals in Language Usage. Cambridge University Press.

[32] Hickey L, Stewart M (2005) Politeness in Europe. Multilingual Matters.

[33] Slugoski BR, Turnbull W (1988) Cruel to be Kind and Kind to be Cruel: Sarcasm, Banter and Social Relations. Journal of Language and Social Psychology 7: 101–121. doi: 10.1177/0261927X8800700202.

[34] Ferguson CA (1976) The structure and use of politeness formulas. Language in Society 5: 137–151. doi: 10.1017/S0047404500006989.

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

[36] Li D. A Contrastive Study of Greetings And Partings In English And Chinese. [Internet] Available from: http://wenku.baidu.com/view/32f0911da8114431b80dd806, Accessed on 1 Nov 2013.

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

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

[39] Polzin TS, Waibel AH (1998) Detecting Emotions in Speech.

[40] Church KW, Hanks P (1990) Word Association Norms, Mutual Information, and Lexicography. Comput Linguist 16: 22–29.

[41] Kuramochi M (2010) Deai no aisatsu kotoba ni okeru koodo to kinō no henka. Gengo to Kouryu 13.

[42] Kuramochi M (2013) Aisatsu kotoba no seisei to hen'yō Meikai University.

[43] Kuramochi M (2013) Generation and Transformation of Aisatsu Words. Meikai University

[44] Lundmark T (2009) Tales of Hi and Bye: Greeting and Parting Rituals around the World. Cambridge University Press.

[45] Scagliotti G, Mujtaba B (2010) Take a bow: culturally preparing expatriates for doing business in Japan. Journal of Comprehensive Research 8: 56–72.

[46] De Mente BL (2003) Kata: The Key to Understanding & Dealing with the Japanese! Tuttle Publishing.

[47] Morris D (1994) Bodytalk: a world guide to gestures. Jonathan Cape.

[48] Do M, Azad P, Asfour T, Dillmann R (2008) Imitation of human motion on a humanoid robot using non-linear optimization. 8th IEEE-RAS International Conference on Humanoid Robots. pp. 545–552.

[49] Carcagnì P et al. (2014) Real-Time Gender Based Behavior System for Human-Robot Interaction. Social Robotics. 8755:74–83.

Gabriele Trovato, Massimiliano Zecca, Martin Do, Ömer Terlemez, Masuko Kuramochi, Alexander Waibel, Tamim Asfour and Atsuo Takanishi:
A Novel Greeting Selection System for a Culture-Adaptive Humanoid Robot

13