

Increasing the Expressivity of Humanoid Robots with Variable Gestural Expressions

Andre Viergutz, Tamara Flemisch, Raimund Dachsel
Interactive Media Lab
Technische Universität Dresden
01062 Dresden, Germany

(andre.viergutz|raimund.dachsel)|@tu-dresden.de, tamara.flemisch@mailbox.tu-dresden.de

ABSTRACT

This work aims at establishing a more variable, and thus attractive, communication with humanoids. For this purpose we developed a method of dynamically expressing emotions and intentions associated with parametrised gestures. First, gestures and possible parameters, which are necessary for the generation of a whole-body gesture set, were analyzed. A gesture's inner and outer expressivity is thereby defined which leads to the differentiation of single gestures and variable gestural expressions. Gestures are subsequently classified into (feedback) categories and related depending on their expressivity. We developed a gesture set consisting of six categories including over 50 variable gestures. As proof of concept, the gesture set has been implemented to allow for an easy and flexible authoring process of gesture-supported communication.

Categories and Subject Descriptors

I.2.9 [Artificial Intelligence]: Robotics—*Operator Interfaces*

Keywords

Non-Verbal Communication; Gestural Interaction

1. INTRODUCTION

The communication as well as interaction with a humanoid robot poses a big challenge to the design of the input- and output possibilities. The interface strongly depends on the embodied attributes of the particular robot. Moreover, a humanoid appearance suggests the increased use of common modes of communication, like speech and gestures.

Beside gesture *input*, which is widely being explored, the articulation of the robot by using whole-body gestures as an *output* and *feedback* modality attracts attention. Multifaceted social signals in co-expression to speech can be recognized fast and effective by human beings. The creation of proper and efficient gestures however is still a great

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

HRI'14, March 3–6, 2014, Bielefeld, Germany.

ACM 978-1-4503-2658-2/14/03.

<http://dx.doi.org/10.1145/2559636.2559840>.



Figure 1: Series of gestures (left to right: "Hand under chin" (expressivity = 0.1), "Scratch head" (0.7) and "Arms bent" (1.0)) increasing the expressivity of *Incomprehension*.

challenge due to personal and cultural imprints. Hence, the implementation of gestures is often realized in an ad hoc and deterministic fashion to fit their special purpose.

We approach this problem by asking 1) What should we communicate by using gestures? and 2) How may a system be designed to easily and effectively select suitable gestures? This work especially addresses the authoring part, for that matter showing a modular and procedural system, which is easy to use while allowing for individual adaptations.

2. RELATED WORK

It has been shown that the usage of gestures by robots, co-expressive to speech, states an effective method to encourage the interaction between humans and robots. Aly and Tapus [1] found that people thereby identify themselves better with a robot, that reveals a more similar behavior. There are several examples of articulating robots. Kobayashi et al. [4] created a museum robot which attracts the visitors attention by head movements and uses deictic gestures to show the way. The expression of emotions for example was realized 2005 on *Leonardo*, a small, sociable robot by Breazeal et al. [2]. Also the humanoid robot Nao, which we used in our project, was used to explore the effects of articulated expressions [5]. To maintain a certain flexibility as well as being able to automatically derive appropriate gestures, certain high-level properties were extracted. Hartmann et al. [3] use the *Overall Activation*, *Spatial-Extent*, *Fluidity*, *Power* and *Repetition* to explain the process of performing a gesture. However, these parameters only refer to a concrete gesture and are not simply applicable to a whole set of expressions. Furthermore, they are restricted in their potential for increase because only a small fraction of the parameter space yields meaningful gestures.

To increase the expressive power beyond a single gesture, we propose *Variable Gestural Expressions*.

3. VARIABLE GESTURAL EXPRESSIONS AND GESTURE SELECTION

Observing the use of gestures during common and everyday communication gave us the insight that many of the used gestures serve specific but often similar purposes. This brought us to the idea to consider whole categories of articulation, not just a single gesture. Therefore, we classified more than 50 commonly used gestures into its expressive meanings and compared them with respect to their basic expressivity. This led us to the definition of six (feedback) categories: *Positive/Acceptance*, *Negative/Denial*, *Incomprehension*, *Neutral/Waiting*, *Introduction/Greeting* and *Ending/Goodbye* of a conversation. To increase a certain intention or expression it is not enough to purely increase the inner parameters of a specific gesture. A timid waving of ones hand in front of the body for example only partially increases its power by just repeating it more often and faster. Simple exaggerations are perceived unnatural or even preposterous. To amplify the attraction of attention as an example of *Introduction* happens by considering the whole arm covering a bigger recognizable area. In fact one could argue, that this kind of augmentation relates to the Spatial Extend but considering the goal to determine a proper gesture this mapping doesn't work in both directions. The adequate movement and the decision about which body parts being involved cannot be stated as a bijective function. The expression of incomprehension is even more complicated (see Figure 1).

However, all these single gestures share a certain *basic* or *outer expressivity* with respect to it's expression category, each limited by an individual *expressivity range* to allow for inner gesture parameter variations. Figure 2 shows an expressivity diagram for the category "*Positive/Acceptance*". The inversion of the mapping from a single gesture to its

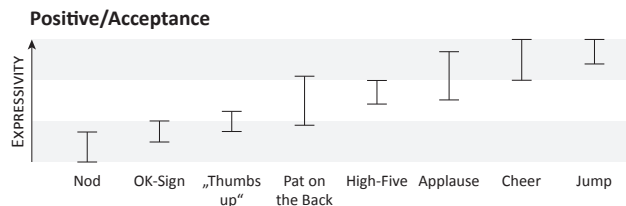


Figure 2: An *expressivity diagram* showing subsequent single gestures for the category "*Positive/Acceptance*"

basic expressivity range leads to a function allowing the selection of one or more suitable gestures by limiting the user only having to set one meta-parameter, the *Expressivity*. In case of multiple choices the system has to decide for a gesture either randomly or depending on the actual context, which is not covered in this work. As mentioned above, each single gesture covers a sub-range of the overall expressivity. Assuming that each of them is parametrically controllable between 0 and 1, the overall expressivity has to be scaled to the relative expressive power.

In addition to determining a suitable gesture, the expressivity also influences inner gesture parameters, like *Frequency*, etc. thereby making inner gesture variations possi-

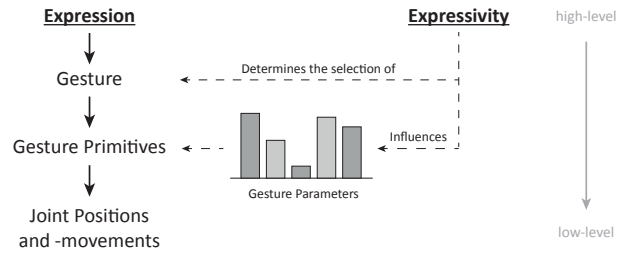


Figure 3: From expressivity to joint positions. The procedural hierarchy of creating a gesture.

ble as to be seen in Figure 3. The mapping of the *Expressivity* to more than one sub-parameter consequently constrains the space of all possible configurations. Particular mapping functions have to be defined, which makes the problem more complex but in turn only meaningful configurations might be created.

4. DISCUSSION AND FUTURE WORK

We proposed a method to reduce the effort of creating as well as selecting suitable gestures to a minimum to be used for enhancing the communication with a robot. Thereby we took care to keep the variety of expressions flexible and dynamic. A library consisting of 26 basic gestures with a total of 50 variations was built for the humanoid robot Nao, which was tested in a small scripted conversation to validate the easy and modular use.

The selection process is currently performed manually by the user. Determining the robots intention and incorporating other contextual information like cultural imprints as well as an individual character of the robot would let our system decide on its own. This could be enhanced in terms of additional meta-parameters. An iterative active learning process for appropriate parameter selection would make the system adaptive and thus more appealing with respect to the correspondence of the user's expectations. A study about the subjective perception will give us further insights into the design and usability of our approach.

5. REFERENCES

- [1] A. Aly and A. Tapus. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. In *Proc. of ACM/IEEE HRI*, pages 325–332, 2013.
- [2] C. Breazeal, C. Kidd, A. Thomaz, G. Hoffman, and M. Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Proc. of IEEE/RSJ IROS*, pages 708–713, 2005.
- [3] B. Hartmann, M. Mancini, and C. Pelachaud. Implementing expressive gesture synthesis for embodied conversational agents. In *GW'05 Proc. of Gesture in Human-Computer Interaction and Simulation*, pages 188–199, 2006.
- [4] Y. Kobayashi, Y. Hoshi, G. Hoshino, T. Kasuya, M. Fueki, and Y. Kuno. Museum guide robot with three communication modes. In *Proc. of IEEE/RSJ IROS*, pages 3224–3229, Sept. 2008.
- [5] R. Meena, K. Jokinen, and G. Wilcock. Integration of gestures and speech in human-robot interaction. In *IEEE CogInfoCom*, pages 673–678, Dec. 2012.