Shaping Robot Gestures to Shape Users' Perception

A Case Study on the Effect of Amplitude and Speed on Godspeed Ratings

Bart Craenen, Amol Deshmukh, Alessandro Vinciarelli and Mary Ellen Foster
University of Glasgow, School of Computing Science
Bart.Craenen@glasgow.ac.uk

ABSTRACT

This paper presents a case study investigating the relationship between the way a robot displays different gestures, and the way these gestures are perceived by its users. The experiments presented show that, at least for some gestures, there is a statistically significant association between changes in amplitude and speed and changes in the responses by the users on the Godspeed questionnaire. Finally this paper presents a description of encountered issues while conducting the case study, a root-cause analysis of those issues, and a discussion of what lessons can be learned from them.

CCS CONCEPTS

• **Human-centered computing → User models**; • **Computer systems organization → Robotic autonomy**; • **Computing methodologies → Computational control theory**;

KEYWORDS

Human-robot interaction, synthetic gestures, perception

ACM Reference Format:
Bart Craenen, Amol Deshmukh, Alessandro Vinciarelli and Mary Ellen Foster. 2018. Shaping Robot Gestures to Shape Users' Perception: A Case Study on the Effect of Amplitude and Speed on Godspeed Ratings. In *Proceedings of HRI (What Could Go Wrong Workshop)*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

This paper presents a case study investigating the relationship between the gestures that a humanoid robot displays and the perception of human interaction partners with the robot. The case study thus investigates the tendency of the users to attribute certain characteristics to a robot rather than others. The main difference to most previous work in this area is that the approach proposed for the case study does not only take into account what the gestures are that the robot displays, but also the way in which the robot displays them. In particular, the case study investigates the association between variations of the *amplitude* and *speed* as well as the variations of the users' perception as measured by the Godspeed questionnaire [1].

In most cases, the main reason for research into gestures is that "gestural expression is intimately involved in acts of spoken linguistic expression" [12], meaning that speech and gestures are processed as a bimodal unit at the neural [9], cognitive [3] and psychological [10] level. In particular, speech and gestures have been shown to mutually enhance each another in order to make an agent more effective in achieving communicative goals [18].

However, the experiments of the case study revolve around the interaction between people and robots in public spaces, and, more specifically, in environments in which the level of acoustic noise tends to be high enough to make it difficult to hear and understand speech by the robot. In situations like this, according to [16, 17], multiple modalities do *not* enhance each other, but rather generate redundancy, by expressing the same message. In this way, the failure of one modality can be compensated for by other modalities. This is why the case study focuses on isolated gestures that do not accompany, or interact with, spoken messages.

For the experiments, 30 independent human observers were asked to watch 45 different gestures performed by an off-the-shelf robot, i.e., *Pepper*, and to complete, for each gesture, the Godspeed questionnaire [1]. The 45 gestures represent 9 variants of 5 core gestures selected from the standard library available with the robot. The variants were obtained by manipulating two parameters; *speed* and *amplitude*. The speed and amplitude parameters were chosen because they are related to energy and spatial extension, respectively; two characteristics that have been shown to play a crucial role in the expressiveness of artificial agents [7]. By analyzing the user responses, the case study investigates whether there is an association between the way a gesture is performed and the perception of the users.

While conducting the case study a number of unexpected issues were encountered. These include situations of a practical nature, but also include unexpected results, as well as responses from the users that did not turn out as expected. This paper not only presents those parts of the case study that went right, but also includes descriptions of the things that did not go as planned or expected, including root-cause analyses of these issues. A discussion on the lessons learned is also included.

The rest of this article is organized as follows: Section 2 describes the experimental approach, Section 3 presents experiments and results. Section 4 then describes a number of issues encountered while conducting the case study. The final Section 5 draws some conclusions.

2 EXPERIMENTAL APPROACH

This section describes the way the gestures adopted in the experiments have been generated (see Section 2.1), and the approach adopted to investigate the relation between users' perception and gestures (see Section 2.2).

What Could Go Wrong Workshop, March 2018, Chicago, IL, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2.1 The Stimuli

This section describes the process aimed at synthesizing the 45 gestures (the *stimuli* hereafter) used in the experiments of this work. The first step is the selection of 5 standard gestures (the *core stimuli* hereafter) available in the library accompanying the *Pepper* robot. The selection targeted gestures that, according to the criteria underlying the taxonomy proposed in [15], are relevant to the scenario addressed in this work, i.e., the interaction between people and robots in public spaces. The names that the robot's manufacturer has given to the selected gestures are as follows [1]:

- Disengaging / Send-away;
- Engaging / Gain attention;
- Pointing / Giving Directions;
- Head-Touching / Disappointment;
- Cheering / Success.

The second step in the process is the synthesis of 9 variants for each of the core stimuli above. Three variants were generated by adopting three different values of the speed λ per core stimulus: 15, 25 and 35 *frames per second* (*fps*), where 25 *fps* is the original speed of the core stimuli. For each of the 15 resulting gestures, another three stimuli were obtained by modifying the differences $\Delta_i(t) = \theta_i(t) - \theta_i(t-1)$, where $\theta_i(t)$ is the angle between the two mechanical elements connected by joint i at frame t. In particular, the values of the $\Delta_i(t)$ were multiplied, for all values of i and t, by a factor α (the *amplitude* hereafter). Three different values of α were adopted, namely 0.50, 0.75 and 1.00. In the first two cases, the result is a dampened version of a core stimulus; in the last case, the $\Delta_i(t)$ are left unchanged.

As a result of the process above, the 9 variants of a given core stimulus correspond to 9 pairs (α, λ), where the pair with $\lambda = 25$ and $\alpha = 1.00$ represents the core stimulus itself.

2.2 Perception Effects Analysis

The first question addressed by the case study is whether users perceive robots that display different gestures differently, and, if so, how the perception changes based on the characteristics of the gestures. During the experiments, the human observers filled in the Godspeed questionnaire [1] after watching each of the 45 stimuli (all observers observed and rated all stimuli). The Godspeed questionnaire is widely accepted as a standard measurement tool for Human Robot Interaction and aims to quantify the following tendencies underlying users' perception:

(1) *Anthropomorphism*: tendency of human users to attribute human characteristics to a robot;
(2) *Animacy*: tendency of human users to consider the robot alive and to attribute intentions to it;
(3) *Likability*: tendency of human users to attribute desirable characteristics to a robot;
(4) *Perceived Intelligence*: tendency of human users to consider intelligent the behavior of a robot;
(5) *Perceived Safety*: tendency of human users to consider safe the interaction with a robot.

Completing the questionnaire results in five scores that measure the tendencies above: the higher the score, the more pronounced the tendency (see [1] for full details). The tendencies were analyzed by defining a χ^2 variable over these scores. The χ^2 variable allows one to test whether the observed distribution of the scores deviates from the uniform distribution to a statistically significant degree. When this is the case, it is possible to say that the Godspeed tendency is more or less pronounced depending on the particular gesture being displayed.

Age Range	18-22	23-25	26-30	31-35	36-40	>40
No. of Subjects	11	6	6	3	1	3

Table 1: Age distribution of the experimental subjects

3 EXPERIMENTS AND RESULTS

The case study involved 30 human observers that were asked to watch the 45 stimuli and, for each of them, to fill out the Godspeed questionnaire (see above). All observers have performed these tasks for all stimuli. The stimuli were presented in random order in three separate sessions (15 stimuli per session). The sessions were held over three consecutive days to limit possible tiredness effects due to the repetition of the tasks over extended periods of time. The 30 observers were split into groups of 3 people, who were asked to participate in the same sessions while still working independently from one another. The observers completed the questionnaires on a tablet while sitting in front of the robot at a distance of roughly 1.5 meters.

The 30 observers were selected randomly from a pool of subjects available at the university where the experiments have been performed: 20 of them are female and 10 are male. The age distributions is available in Table 1, and the participants reported a range of different ethnic and national origins. Only 3 observers had interacted with a robot before participating in the experiments. Each participant received a payment corresponding to the minimum legal hourly wage in the country where the experiments have been performed.

3.1 Gestures and Perception

Table 2 shows the cases in which the distribution of the Godspeed scores across the multiple variants of the same core stimulus deviates, to a statistically significant extent, from the uniform distribution. The table also shows whether increasing amplitude and speed of a gesture corresponds to higher or lower Godspeed scores. A deviation from the uniform distribution is considered statistically significant when a χ^2 test results into a p-value lower than 0.05. The *False Discovery Rate* (FDR) correction [2] was applied to tackle the multiple comparisons problem.

For the Disengaging gesture, a significant effect was found for Likability and Perceived Safety. In the former case, the scores tend to decrease when α and λ increase, while in the latter case the scores tend to increase when α and λ decrease. The possible explanation

[1]The animations associated to the core stimuli are available on the version 1.6B of Pepper in the following directories:
"animations/Stand/Gestures/No_3" (Disengaging),
"animations/Stand/Gestures/Hey_2" (Engaging),
"animations/Stand/Emotions/Negative/Hurt_1" (Head-Touching) ,
"animations/Stand/Gestures/Far_3" (Pointing) and
"animations/Stand/Emotions/Positive/Happy_1" (Cheering).

	Ant		Ani		Lik		Int		Saf	
Core Stimulus	α	λ	α	λ	α	λ	α	λ	α	λ
Engaging	↑	↑	↑	↑	↑	↑				
Disengaging					↓	↓			↓	↓
Pointing										
Head-Touching			↑	↑						
Cheering			↑	↑						

Table 2: The symbols "↑" and "↓" account for statistically significant effects. The symbol "↑" means that increasing amplitude or speed corresponds to observing higher Godspeed scores. The symbol "↓" means that decreasing amplitude or speed corresponds to observing lower Godspeed scores. Empty cells correspond to cases in which no statistically significant effects have been observed.

behind the Likability effects is that this gesture aims to increase the physical distance between the robot and its users. Given that physical and social distances have been shown to be equivalent [11], increasing the energy of the gesture might appear to be an attempt by the robot to push people towards distances that, according to proxemic theories [6], correspond to less friendly and more formal relationships. With respect to Perceived Safety, the probable explanation is that slower movements (lower λ) that do not extend far from the robot's body (lower α) appear less likely to harm the users.

In the case of the Engaging gesture, statistically significant effects have been observed for Anthropomorphism, Animacy and Likability. In all three cases, increasing the amplitude and speed corresponds to higher Godspeed scores. In the case of Anthropomorphism, one possible explanation is that the human brain has been shown to be more anthropomorphic when synthetic movements are more similar to those displayed by humans [5]. The possible explanation for the Animacy effects is that higher speed and amplitude result into higher energy and motor activation, two factors that play a crucial role when it comes to consider an agent alive or lively [1]. Finally, the increase in Likability scores is likely to depend on the correlation between Anthropomorphism and positive judgments about the robots that have been observed earlier in the literature [20].

Overall, the three effects observed for the Engaging gesture are an advantage in those scenarios in which the robot is expected to pro-actively initiate an interaction with the users: in particular, the effects indicate a possible mechanism for generating more positive perceptions, a prerequisite towards successful interactions with machines that display human-like behavior (see, e.g., [14]).

No statistically significant effect was observed for the Pointing gesture. A possible explanation is that deictic gestures such as pointing are meant to convey information about spatial knowledge [8] and not about the social and psychological phenomena underlying the items of the Godspeed questionnaire [1].

Finally, both the Head-Touching and Cheering gestures show significant effects on the Animacy scale. The probable reason is that both gestures, when displayed by people, tend to convey information about one's inner state. In particular, Head-Touching is typically associated with a situation of confusion [13, 19] while Cheering tends to be displayed as a sign of success, satisfaction, and accomplishment [4]. This means that a robot displaying these two gestures can elicit the attribution of the same inner states and, ultimately, of Animacy, defined as the very property of being alive [1].

For both Head-Touching and Cheering, the Animacy scores tend to increase when both α and λ increase. In the case of α, the probable reason is that lowering this parameter leads to gestures that have a morphology different from the core stimulus and, hence, fail in conveying the same impression. In the case of λ, the probable reason is that movements have been shown to play a crucial role in the attribution of Animacy, the very difference between animate beings and inanimate objects [1].

4 ENCOUNTERED ISSUES

While conducting the case study, a number of unexpected issues were encountered. Instead of only presenting the parts of the case study that went right, this paper also includes descriptions of the things that did not go as planned or expected.

One issues that occurred is of a practical nature. Using off-the-shelf robots often also means being, in some way, tied to the manufacturer's supplied software and operating system. In the case of Pepper this is no different. One notable issue that occurred early on in the experiment is that the robot would, unexpectedly, and unprovoked, start tracking one of the three observers. This did not happen during initial tests of the experiment, and had thus went unnoticed.

Early attempts at removing this unwanted feature failed because it was caused by issues in the supplied operating system. Since this behavior may affect the perception of the robot this issue needed to be dealt with however, and was only rectified by the expedient of covering up the depth camera of pepper. Since the issue was rectified relatively quickly, the effect on the results of the experiment turned out to be negligible. Lessons learned from this issue include conducting more exhaustive initial tests of the experiment, and to never underestimate the expediency of practical solutions for practical problems.

Other issues were caused by errors made in the implementation and/or design of the experiments, in particular of the questionnaire forms. One, quickly and easily rectified error, was an error make in the Godspeed questionnaire, where instead of the proscribed five scale Likert scale [1], an erroneous six point scale was used. By redoing some sessions, this issue had no effect on the end results.

Another, more insidious mistake was made in the design of the other questionnaires the observers were tasked to fill out. Unreported in this paper, the case study also included a part where the observers were tasked to assess the meaning of the gestures. For this 10 possible interpretations were presented. The mistake was to ask the observers to rate all 10 interpretations on a Likert scale, instead of having the observers pick the best out of the interpretations. This necessitated, afterwards, the use of some advanced statistical analysis (relative entropy etc.), the results of which proved to be somewhat difficult to interpret, and represent. Lessons learned from these two issues include nothing more profound that to thoroughly check and test all questionnaires before use, and, to always keep in mind, beforehand, what analysis will be required after the answers have been collated.

A final issue to discuss is that in the case study, the expectations of the participants with respect to the robot, i.e., Pepper, did not always correspond with the expectations that the conductors of the case study had of these expectations. Given that Pepper is a well-publicized social robot, the conductors of the experiment expected that a fair number of participants would have had prior, and positive, experience or knowledge of Pepper. In actuality, only two participants had prior experience with robots, and, surprisingly, that experience turned out to be rather negative towards Pepper, and more specifically, its capabilities. Many other participants expressed interest in robots, and/or Pepper (some enthusiastically so), but some also, afterwards, expressed disappointment in the case study, pointing out the limited personal interaction they were allowed to have with the robot. There is no evidence that this had any effect on the results of the case study, but, especially for the more enthusiastic participants, one lesson learned is to pay sufficient attention to managing expectations when advertising for these types of case studies.

5 CONCLUSIONS

This study presents a case study on the relation between the way a gesture is performed, and the perception of the users. The results show that, at least in some cases, there is an association between the speed and amplitude of a gesture (two parameters that account for energy and spatial extension) and the scores on the Godspeed questionnaire [1].

Overall, the coherent picture that emerges is that, for gestures expected to achieve a social goal (Engaging and Disengaging) effects were found on the Godspeed dimensions that better account for social aspects of Human Robot Interaction, namely Anthropomorphism (the tendency to attribute human characteristics to the robot) and Likability (the tendency to attribute desirable characteristics to the robot). Similarly, gestures designed to simulate an “inner state” (Head-Touching and Cheering) show effects on Animacy, the Godspeed dimension that captures the tendency to consider the robot alive and, hence, capable of experiencing the world.

Finally, there are no effects for Pointing, which, unlike the other stimuli used in the experiments, aims to share knowledge about the environment more than to convey information about the robot properties embodied in the Godspeed questionnaire.

The above suggests that the stimuli have been designed correctly and, most importantly, it shows that the Godspeed scores tend to be different for different values of amplitude and speed. The main implication of such an observation is that, for a social robot, it is not sufficient to decide what gestures a robot should display during an interaction, but also how the gestures are performed. In particular, the same gesture should be displayed with different amplitude and speed depending on the target impression to convey on the tendencies underlying the Godspeed scores.

The paper also includes a section on unexpected issues that occurred while conducting the case study. These include issues of a more practical nature to do conducting experiments with robots, issues with the design of the experience and the questionnaires, and the issue of (managing) expectation in robotics research. Root-cause analyses are included, and lessons learned from the issues highlighted are described.

Future work will aim at investigating how the findings of this study can change when the gestures are accompanied by speech, as in the most frequent case in everyday interactions [12, 18, 21]. Furthermore, future work will investigate the relation between gestures and other characteristics that users can attribute to the robot such as, e.g., the Big-Five personality traits.

ACKNOWLEDGMENTS

This research has been partially funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 688147 (MuMMER, mummer-project.eu). It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

REFERENCES

[1] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (2009), 71–81.
[2] Y. Benjamini and Y. Hochberg. 1995. Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* (1995), 289–300.
[3] J.P. de Ruiter. 2000. The production of gesture and speech. In *Language and Gesture*, D. McNeill (Ed.). Cambridge University Press, 284–311.
[4] C. Gardair. 2013. *Assembling Audiences.* Ph.D. Dissertation. Queen Mary University of London.
[5] V. Gazzola, G. Rizzolatti, B. Wicker, and C. Keysers. 2007. The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage* 35, 4 (2007), 1674–1684.
[6] E. Hall. 1959. *The silent language.* Doubleday.
[7] B. Hartmann, M. Mancini, and C. Pelachaud. 2005. Implementing expressive gesture synthesis for embodied conversational agents. In *Proceedings of International Gesture Workshop.* 188–199.
[8] J. Haviland. 2000. Pointing, gesture spaces, and mental maps. In *Language and Gesture*, D. McNeill (Ed.). Cambridge University Press, 13–46.
[9] S.D. Kelly, C. Kravitz, and M. Hopkins. 2004. Neural correlates of bimodal speech and gesture comprehension. *Brain and language* 89, 1 (2004), 253–260.
[10] S.D.D Kelly, A. Özyürek, and E. Maris. 2010. Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science* 21, 2 (2010), 260–267.
[11] A. Kendon. 1990. *Conducting Interaction.* Cambridge University Press.
[12] A. Kendon. 2000. Language and gesture: unity or duality? In *Language and Gesture*, D. McNeill (Ed.). Cambridge University Press, 47–63.
[13] M.L. Knapp and J.A. Hall. 1972. *Nonverbal Communication in Human Interaction.* Harcourt Brace College Publishers.
[14] C. Nass and S. Brave. 2005. *Wired for speech: How voice activates and advances the human-computer relationship.* MIT Press.
[15] C.L. Nehaniv, K. Dautenhahn, J. Kubacki, M. Haegele, C. Parlitz, and R. Alami. 2005. A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on.* IEEE, 371–377.
[16] S.R. Partan and P. Marler. 1999. Communication goes multimodal. *Science* 283, 5406 (1999), 1272–1273.
[17] S.R. Partan and P. Marler. 2005. Issues in the Classification of Multimodal Communication Signals. *The American Naturalist* 166, 2 (2005), 231–245.
[18] I. Poggi. 2007. *Mind, hands, face and body. A goal and belief view of multimodal communication.* Weidler.
[19] V.P. Richmond, J.C. McCroskey, and S.K. Payne. 1991. *Nonverbal behavior in interpersonal relations.* Prentice Hall.
[20] M. Salem, F. Eyssel, K. Rohlfing, S. Kopp, and F. Joublin. 2013. To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics* 5, 3 (2013), 313–323.
[21] M. Salem, S. Kopp, I. Wachsmuth, K. Rohlfing, and F. Joublin. 2012. Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics* 4, 2 (2012), 201–217.