An Examination of Whether People Prefer Agents Whose Gestures Mimic Their Own

Pengcheng Luo[1], Victor Ng-Thow-Hing[2] and Michael Neff[1]

[1] UC Davis, USA, {`pcluo`|`neff`}`@ucdavis.edu`
[2] Honda Research Institute USA, Inc. `vng@honda-ri.com`

Abstract. Do people prefer gestures that are similar to their own? There is evidence that in conversation, people will tend to adopt the postures, gestures and mannerisms of their interaction partners [1]. This mirroring, sometimes called the "chameleon effect", is associated with affiliation, rapport and liking. It may be that a useful way to build rapport in human-agent/robot interaction is to have the agent/robot perform gestures similar to the human. As a step towards that, this study explores if people prefer gestures similar to their own over gestures similar to those of other people. Participants were asked to evaluate a series of agent motions, some of which mimic their own gestures, and rate their preference. A second study first showed participants videos of their own gesturing to see if self-awareness would impact their preference. Different scenarios for soliciting gesture behavior were also explored. Evidence suggests people do have some preference for motions similar to their own, but self-awareness has no effect.

Key words: Human Agent Interaction, Gesture Mimicry

1 Introduction

With the emergence of more intelligent agents and robots, there is a growing need for improved human-agent/robot interaction. Along with speech, non-verbal behavior plays an important role in communication and interaction. Numerous researchers have been working on developing gestures for agents [1–3] and robots [4, 5]. However, how an agent/robot can most effectively use gestures to interact with people is not well studied. Mimicry is one important phenomenon that is found in human-human interaction. Called "chameleon effect" in some literature, previous researchers have found that mimicry can increase liking among interaction partners [6, 7]. This phenomenon suggests that people may like gestures more similar to their own, and by extrapolation, since humanoid agents/robots share a similar morphology to people, having the agent copy the user's gestures may increase the user's comfort. We conducted a set of experiments to see whether people prefer virtual agents that perform gestures similar to their own. Before the second round of experiments, participants were shown recorded video segments of their movements to see if the increased self-awareness this generated affected their ratings. These studies provide evidence for the efficacy of potential future systems that perform mimicry in real time.

Our contributions lie in several parts: first, we conducted a pilot study on people's preference for agents that gesture like them. The experiments showed some evidence supporting that 1) People appear to show preference for agents that mimic their own gestures, although it may not be their favorite one. A post-experiment analysis raised the possibility that whether people prefer agents that mimic their own motions might be related to people's personality, but this requires further investigation. 2) Self-awareness does not affect people's preference. Secondly, several methods for soliciting gesture behavior have been tested and compared. Analysis shows the type of gestures people perform is affected by content of the prompt they are given.

The structure of the paper is as follows. Section 2 provides relevant literature. Section 3 describes an overview of the experiment design. In Sections 4 and 5, we will discuss the experiment preparation and results. We will conclude with our results a discussion of future directions.

2 Background

Nonverbal behavior of embodied agents and robots, especially gestures, has a strong impact on communication. Krämer et al. [8] found that self-touching gestures have positive effects on user' evaluations. The agent was rated as more natural, warmhearted, agile and committed when presenting self-touching gestures. Salem et al. [5] found that a robot is evaluated more positively when non-verbal behaviors such as hand and arm gestures accompany speech. Neff et al. [9, 10] conducted experiments to understand how the Big Five traits of emotional stability and extraversion correlate with changes in verbal and nonverbal behavior. The perception of these traits is varied by adjusting a virtual agent's gesture rate, whether self-adaptors are present and movement style parameters.

The "chameleon effect" impacts interpersonal interactions. Lakin and Chartrand [7] showed that mimicry can be used to build liking among interlocutors. Chartrand and Bargh [6] found that mimicry facilitates the smoothness of interactions and increases liking between interaction partners. Lee [11] found mimicry to be predictive of liking between participants instead of trust. Several behaviors have been observed to be mimicked in previous literature. Chartrand and Bargh [6] found rubbing one's face or shaking one's foot is mimicked. Cappella and Planalp [12] found that people have a tendency to mimic the way they speak, matching features such as rhythm and pauses. Gestural mimicry [13–15], the recurrence of gestural features across speakers, indicates one speaker's gesture is influencing the others. This inspired our work on examining gestural mimicry as a mechanism for human agent interaction.

For human-agent/robot interaction, mimicry has been used before in several places. The effect of agent mimicry, however, is still an open question. Bailenson and Yee [16] studied the effect of head mimicry using an embodied artificial agent. Their results showed that mimicking agents were more persuasive and received more positive trait ratings than non-mimickers, despite participants inability to explicitly detect the mimicry. Gratch and his colleagues [17] designed a listening agent that would try to create rapport by tying listening feedback

to shallow features of a speaker's voice and bodily movements. In contrast to that research, we are looking at communicative gestures for a speaking agent. Kopp [18] designed a framework for human agent interaction by transferring several coordination mechanisms, such as mimicry, alignment, and synchrony, from face-face interaction to human agent interaction. The primary focus of this paper is to study people's attitude towards agents that gesture like them.

3   Overview

The objective of this research is to study people's attitude towards an agent that copies their own gestures. With this objective in mind, we have two hypotheses:

- Hypothesis1: People prefer gestures that are similar to their own.
- Hypothesis2: Self awareness affects peoples preference for agent motions.

While a real time mimicry system is under development, we implemented "gesture mimicry" by copying gestures performed by subjects during test sessions and applying these to a virtual agent afterwards. To achieve this, we first collected subjects' motion while they discussed different topics. This yielded a gesture database including gesture and key word information for each person. To generate new motion for a virtual agent, we first constructed new text with key words that appeared in the previous topics. Based on the mapping between gestures and key words, we selected gesture motion that is associated with key words from the database. More details can be found in section 4.1.

A pre-test was performed to confirm that the agent gestures accurately replicated the gestures of each subject. An agent video was made based on each subject's data. Side-by-side videos were then made of every combination of agent motion with videos of the original subject motion. These were posted to Amazon Mechanical Turk and subjects were asked to rate the similarity between the original videos and the agent motion. This ensured that the generated agent motion was able to demonstrate mimicry by confirming that the copied motions were the best match for the originals (Section 4.3). The newly generated virtual agents were then shown to the original subjects to see if they prefer the virtual agent that has similar motion to their own.

4   Experiment Preprocess

4.1   Elicitation of Motion Examples

Our first task was to build a motion database that will allow us to generate novel virtual agent utterances that make appropriate use of the captured gestures. This requires satisfying several objectives: 1) we need to collect enough gesture samples for each person, 2) participants should feel comfortable so they perform natural gestures, 3) for the easiness of synthesizing new motions based on same text, we need solicit motions from different people using similar key words. It remains unclear how to best solicit such natural gestures, so we experimented with several approaches. We suspect that speech content is an important factor

that affects gesture quality, therefore we designed 5 tasks based on different content constraints.

1. FiC: Fixed Content with emphasis specified for several key words. Participants are told to speak a short piece of text with emphasis on specified key words.
2. FrC: Free Content. We let participants tell a story freely related to "catching a fish".
3. CV: Constrained by Video. We showed a piece of short video in which the robot Asimo is pouring a drink into a cup and let people describe what happened in the video.
4. CA: Constrained by Audio. We first broadcast a short audio clip about the tale of the "three little pigs" and then ask people to rephrase the story.
5. CI: Constrained by Image. We show people one image and let them describe the scenario in the image.

Eight regular university students participated in this long term research project. Each students was paid $25 for participating in the whole project. During the capture session, we used Kinect to capture 3-d motion and a video camera for image and voice. The motion data was then segmented into gesture sequences using Anvil software [19]. Segmented Kinect data were stored in a database for future use.

We looked at several motion features in the gesture sequences as shown in Table 1. The data shows that task CV (constrained by video) might be the most efficient and stable way to produce motion over different subjects as it has highest average MPD (duration of movement/duration of complete video sequence) and lowest std of MPD. FiC would generate a consistent number of gestures as it has lowest std in terms of number of gestures. CV has second lowest variance in terms of NOG.

Table 1. Motion Statistics. NOG: number of gestures; MPD: duration of movement/duration of complete video sequence; MD: sequence duration(seconds).

experiment type	mean of NOG	std of NOG	mean of MPD	std of MPD	mean of MD
FiC	8.75	1.49	0.43	0.15	15.8
FrC	14.5	11.65	0.75	0.16	51.9
CV	10.5	5.15	0.77	0.10	39.6
CA	34.75	11.55	0.49	0.13	160.4
CI	14.375	9.62	0.66	0.21	53.9

In [20], McNeill found a phenomenon - "the saliency of dimension" - when people describe cartoons in different contexts. Iconic gestures dominate in narrative contexts (promoting the development of the story) while metaphoric gestures predominate in extra-narrative clauses (description of the setting and characters, summaries, etc.). We examined the gesture types performed with in different tasks (Table 2) and found that iconic gestures dominate in CV. The reason

might be that people tend to copy the robot's action while describing the video. If the text is fixed, people tend to use more beats. In a post-experiment survey, 3 of 8 participants mention that task FiC is difficult as they need to read the text and act it out. This could partially explain why there are more beats in this task. No one said that task CV was difficult. 6 of 8 participants agree that task CI, describing a picture, is the easiest one. Therefore we see a relatively balanced number of different gesture types, except emblems which are consistently rare in our motions. When the task is easy, participants might be more expressive using different kinds of gestures.

Table 2. Motion Dimension.

	Emblems	Iconic	Metaphoric	Deictic	Beat	Total Number
FiC	0.00%	15.87%	15.87%	19.05%	**49.21%**	63
FrC	4.07%	48.78%	13.01%	4.07%	30.08%	123
CV	2.25%	**69.66%**	6.74%	10.11%	11.24%	89
CA	2.92%	22.99%	19.71%	8.39%	45.99%	274
CI	2.56%	38.46%	21.37%	16.24%	21.37%	117

4.2 New Motion Generation

It is necessary to have a common script for the virtual agent in order to generate our stimuli for comparison, but people made different speeches during the capture session. To develop a script, we first picked several typical words from FiC and CI based on two considerations: 1) the easiness of finding similar words across different people, 2) the easiness of detecting gesture shape differences given similar key words. We then created two new pieces of text for generating new motions using these key words. Voice was generated using NaturalReader [21], a text to speech software. To limit gender bias, we displayed the motion on a virtual wooden model.

The two new texts used in our generation tests are listed as follows (italicized words are key words that received gestures):

1. "We will *cut* taxes and *raise* the standard of living in this country. For *all* of you hurting out there, I *feel* your pain. "
2. "In the *picture*, there is a *big* house in the upper corner with thousands of *balloons*. A *rope* is coming out of it."

The captured motions in the database were segmented and labeled with associated key words. They were retargeted to the virtual character before motion reuse. Post-processing tools were applied to make sure the motions look similar to the original motion and this similarity was ensured through the pre-test discussed in Section 4.2. To generate personalized, new motions, we align motion segments in a personalized database with their associated key words in the text. During alignment, we make sure the end of the stroke is aligned with the end of key word audio. Motion segments are connected to create continuous sequences. In total we have produced 8 personalized agent motion clips for each utterance.
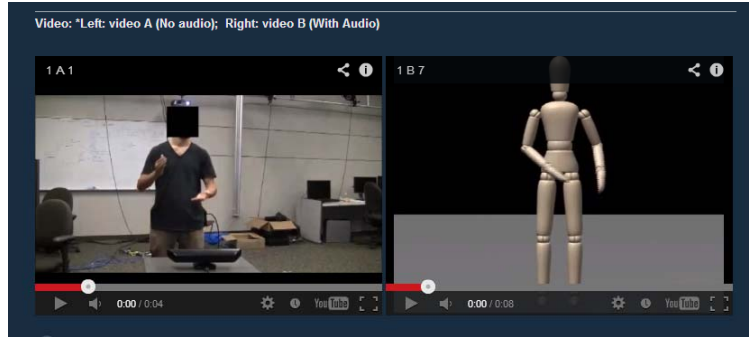
4.3 Similarity Evaluation

Fig. 1. The snapshot of similarity evaluation experiment.

We want to ensure that there is clear "mimicry" between a subject's original performance and the sequence designed to match it. It is not an easy task to quantify the similarity between a pair of motion sequences, thus we employ a perceptual study on Amazon Mechanical Turk to evaluate the similarity between the original video and new motion. We showed participants motions side by side: left, original video (OV) and right, the generated motion (GM)(Fig. 1). After viewing the motion, they are asked to provide a rating for each of the following 5 prompts on a 5-point Likert-scale "Strongly disagree/Disagree/Neutral/Agree/Strongly agree" with values from 2 to -2.

- Character in video B seems to have the same personality as the person in video A.
- Character in video B appears to have the same natural rhythms (timing) as the person in video A.
- Character in video B is trying to copy the person in video A.
- Character in video B is different in attitude from the person in video A.
- Character in video B has a similar expressiveness to video A.

The average of the five questions (note the rating of the fourth item is reversed) will be used to indicate the similarity level between the videos and generated motion sequence.

Let (OV_i, GM_i) define a pair of motion sequences. The new motion GM_i is generated from original motion captured along with OV_i for subject i. Since there are 8 motion pairs, every participant needs to compare OV_i with $\{GM_k | k = 1, 2, \cdots 8\}$. For OV, we only showed video clips that are associated with the key words used in GM. In order to minimize potentially contaminating effects, we blacked out the face and turned off the voice. The motion clips are connected in a cross-dissolve format as we want to let people focus on the gestures while avoiding the abruptness of sudden cuts. To verify the similarity of 16 generated

agent videos, we designed 16 perceptual experiment with 8 pairs of video in each. 10-15 participants were collected for each experiment and each was paid \$0.5/test. In order to collect high quality surveys, we only recruited "master workers"(qualified participants on Amazon Mechanical Turk) and reviewed the results before approving the payment. The results showed that the average similarity rating of the pair (OV_i, GM_i) is always the highest compared with other pairs (OV_i, GM_j) where $j \neq i$. This means that we have correctly produced new motions that have noticeable mimicry. The average ratings for each clip are shown in the following Table 3.

Table 3. Average similarity rating based on 5 questions above (first utterance).

original vs generated motion	1	2	3	4	5	6	7	8
1	**0.36**	0.00	-0.18	-0.45	-0.45	-0.27	0.00	-0.18
2	-0.43	**0.29**	-0.79	-0.50	-0.79	-0.50	-0.14	0.07
3	0.00	-0.75	**0.33**	-0.58	-0.50	-0.58	-0.08	-0.08
4	-0.43	-0.57	-0.79	**0.14**	-0.71	-0.43	-0.57	0.07
5	-0.22	-0.44	-0.56	-0.78	**0.78**	0.11	-0.67	-0.22
6	-0.36	-0.45	-0.55	-0.55	0.27	**0.45**	-0.45	-0.18
7	-0.17	-0.25	-0.08	-0.67	-0.58	-0.42	**0.42**	-0.25
8	-1.00	-0.40	-0.80	-0.70	-0.30	-0.20	-0.50	**0.60**

5 Experiment and Result Analysis

5.1 Overview

We designed two experiments to test our two hypotheses: Hypothesis1: people prefer the gestures that are similar to their own; Hypothesis2: self awareness affects peoples preference for agent motions.

The first experiment tests whether people prefer the motion that is similar to their own. To achieve this goal, we show two groups of generated agent motions to participants whose motions were recorded before. Each group has 8 videos for one of the utterances which were played in random order (only one animation is based on the participant's data). In order to achieve a more accurate rating on affinity, we designed questions along seven different character traits (persuasive/likable/charismatic/excited/competent/trustworthy/friendly) with ratings form 1 to 9 (least to most) in the format "How persuasive does this character seem?". A post analysis shows that these characteristics are highly correlated (the Pearson correlation coefficients range from 0.69 to 1.0). Therefore we used the average rating as an indicator of their affinity rate.

The second experiment examines whether self-awareness will affect people's preference for self-similar motion. To evaluate this, after showing the first set of motions, we showed participants own source videos in the format used in section 4.3. After that, we replayed the generated motions, again in random order, to see if there were any differences in ratings.

The original participants worked on this experiment 3 months after we recorded their motions. We also recorded the time duration that participants spent on each page, with which we can detect invalid data (if the time was too short, it is unlikely that participants were honestly evaluating the clips). 8 students participated in our research and 6 of them had valid data.

5.2 Experiment 1: Preference for Self-similar Motion

To test whether people prefer the motion that is similar to their own, we first rank the ratings of all the agent motions that they watched. If the rank of the agent motion that mimics their own motion ranked 1st, then it indicates that people prefer mimicry motions.

As we only have limited data and the distribution of ranks are not clear, we used a nonparametric test, the sign test [22]. The p-value of the test is 0.92 (n=24). This indicates that the median of the targeted clip rank is greater than 1 and the agent motion that has mimicry is not their favorite one. However if we relax the hypothesis to compare median rank to 4 (as we only have 8 agent motions), p-value < 0.05. The result indicates that the agent motion that has mimicry is on their preferable motion lists.

We also compare the rating of GM_i and the average rating of GM_j where $j \neq i$ without self-awareness. The results show that subject 2 and subject 5 particularly do not like their own motion, but other subjects show preference to their own motion compared with other averaged ones (Fig. 2). A follow up big 5 personality test [23, 24] shows that these two subjects are less extroverted as their test scores are in percentile 24% and 8% in a data groups with 711 users examined in [25]. This might suggest that whether people prefer mimicry is related to individual personality which supports von der Pütten et al.'s findings [26]: participants' personality traits influence their subjective feeling and evaluation of virtual characters.

5.3 Experiment 2, the Factor of Self-awareness

The next experiment examined the factor self-awareness. Due to the small data size, we used signed rank test [22] to see if there is any difference on rating of GM_i before and after showing subject i his own videos. The signed rank test shows that the self-awareness effect is not significant (p=0.46). We also used the t-test to look at the real rating of the targeted video before and after watching their own videos. t - test shows that there is no significant difference (p = 0.88).

6 Conclusion and Discussion

This paper describes a pilot study on whether people prefer gestures that look like their own, providing a potential way of improving human-agent interaction. Through experiments we found that 1) people appear to show a preference for gestures that copy their own but they are not their favorite choice, 2) self-awareness does not affect people's preference for self-similar gestures.

We found several interesting phenomena by analyzing the data which shed light on interesting future directions. 1) In our study, two subjects did not prefer mimicked gestures, but four did. This suggests that whether people prefer
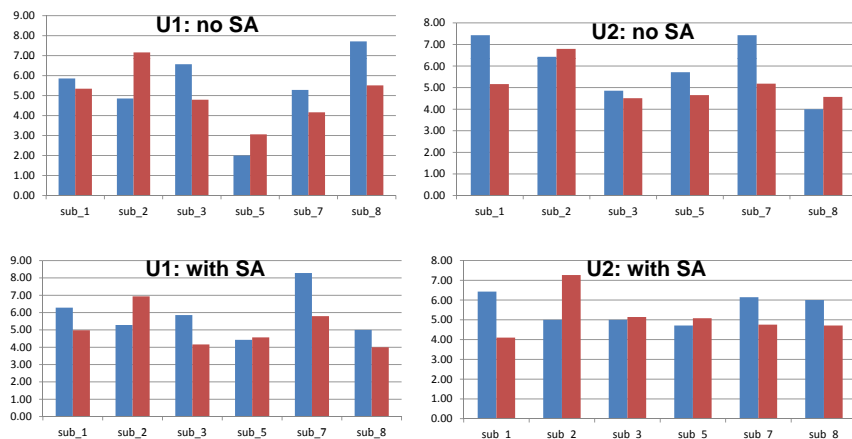
Fig. 2. Horizontal axis: subject i; Vertical axis: rating; Blue: rating of GM_i; Red: average rating of other motions; SA: self-awareness; U1: first utterance; U2: second utterance.

mimicry may be quite personality/person dependent. 2) Gesture types seem to be related to content. Task CV seems to have a high chance to solicit iconic gesture as participants tend to copy actions in the video; task CI is able to solicit gestures of balanced distribution in different categories. Guidelines on selecting appropriate content for gesture productions are as follows: if iconic gestures are the highest priority, describing a video with actions might be a good task; asking people to act out a piece of text would probably induce more beat gestures; describing an image would be likely to generate gesture types of balanced distribution.

References

1. Cassell, J., Vilhjálmsson, H.H., Bickmore, T.: Beat: the behavior expression animation toolkit. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM (2001) 477–486
2. Kopp, S., Wachsmuth, I.: Synthesizing multimodal utterances for conversational agents. Computer animation and virtual worlds **15**(1) (2004) 39–52
3. Neff, M., Kipp, M., Albrecht, I., Seidel, H.P.: Gesture modeling and animation based on a probabilistic re-creation of speaker style. ACM Transactions on Graphics (TOG) **27**(1) (2008) 5
4. Ng-Thow-Hing, V., Luo, P., Okita, S.: Synchronized gesture and speech production for humanoid robots. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE (2010) 4617–4624
5. Salem, M., Rohlfing, K., Kopp, S., Joublin, F.: A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction. In: RO-MAN, 2011, IEEE (2011) 247–252

6. Chartrand, T., Bargh, J.: The chameleon effect: The perception–behavior link and social interaction. Journal of personality and social psychology **76**(6) (1999) 893
7. Lakin, J., Chartrand, T.: Using nonconscious behavioral mimicry to create affiliation and rapport. Psychological Science **14**(4) (2003) 334–339
8. Krämer, N.C., Simons, N., Kopp, S.: The effects of an embodied conversational agents nonverbal behavior on users evaluation and behavioral mimicry. In: Intelligent virtual agents, Springer (2007) 238–251
9. Neff, M., Toothman, N., Bowmani, R., Tree, J.E.F., Walker, M.A.: Dont scratch! self-adaptors reflect emotional stability. In: Intelligent Virtual Agents, Springer (2011) 398–411
10. Neff, M., Wang, Y., Abbott, R., Walker, M.: Evaluating the effect of gesture and language on personality perception in conversational agents. In: Intelligent Virtual Agents, Springer (2010) 222–235
11. Lee, J.: Modeling the dynamics of nonverbal behavior on interpersonal trust for human-robot interactions. PhD thesis, Massachusetts Institute of Technology (2011)
12. Cappella, J.N., Planalp, S.: Talk and silence sequences in informal conversations iii: Interspeaker influence. Human Communication Research **7**(2) (1981) 117–132
13. Kimbara, I.: On gestural mimicry. Gesture **6**(1) (2006) 39–61
14. Parrill, F., Kimbara, I.: Seeing and hearing double: The influence of mimicry in speech and gesture on observers. Journal of Nonverbal Behavior **30**(4) (2006) 157–166
15. Kimbara, I.: Gesture form convergence in joint description. Journal of Nonverbal Behavior **32**(2) (2008) 123–131
16. Bailenson, J.N., Yee, N.: Digital chameleons automatic assimilation of nonverbal gestures in immersive virtual environments. Psychological science **16**(10) (2005) 814–819
17. Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R., Morency, L.: Virtual rapport. In: Intelligent Virtual Agents, Springer (2006) 14–27
18. Kopp, S.: Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. Speech Communication **52**(6) (2010) 587–597
19. Heloir, A., Neff, M., Kipp, M.: Exploiting motion capture for virtual human animation. In: Proceedings of the Workshop Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality at LREC-2010. (2010)
20. McNeill, D.: Gesture and thought. University of Chicago Press (2008)
21. NaturalSoft: http://www.naturalreaders.com/.
22. Wolfe, D.A., Hollander, M.: Nonparametric statistical methods. Nonparametric statistical methods (1973)
23. John, O.P., Naumann, L.P., Soto, C.J.: Paradigm shift to the integrative big five trait taxonomy. Handbook of personality: Theory and research **3** (2008) 114–158
24. John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventoryversions 4a and 54. Berkeley: University of California, Berkeley, Institute of Personality and Social Research (1991)
25. Benet-Martínez, V., John, O.P., et al.: Los cinco grandes across cultures and ethnic groups: Multitrait multimethod analyses of the big five in spanish and english. Journal of personality and social psychology **75** (1998) 729–750
26. von der Pütten, A.M., Krämer, N.C., Gratch, J.: How our personality shapes our interactions with virtual characters-implications for research and development. In: Intelligent Virtual Agents, Springer (2010) 208–221