Backchannel Head Nods in Danish First Meeting Encounters with a Humanoid Robot: The Role of Physical Embodiment

Anders Krogsager[1], Nicolaj Segato[1], and Matthias Rehm[2]

[1] School of Information and Communication Technology, Aalborg University, 9000 Aalborg, Denmark
{akrogs09,npeder09}@student.aau.dk

[2] Faculty of Engineering and Science, Aalborg University, 9000 Aalborg, Denmark
matthias@create.aau.dk

Abstract. Head nods have been shown to play an important role for communication management in human communication, e.g. as a non-verbal feedback signal from the listener. Based on a study with virtual agents, which showed that the use of head nods helps eliciting more verbal input from the user, we investigate the use of head nods in communications between a user and a humanoid robot (Nao) that they meet for the first time. Contrary to the virtual agent case, the robot elicited less talking from the user when it was using head nods as a feedback signal. A follow-up experiment revealed that the physical embodiment of the robot had a huge impact on the users' behavior in the first encounters.

Keywords: Culture-aware robots, backchannels, feedback, physical embodiment.

1 Introduction

Robots have begun to move from restricted environments that are specially designed for them into public and semi-public spaces where they are envisioned to interact in a socially acceptable manner with users. Head nods have been shown to play an important role for communication management in human communication, e.g. as a non-verbal feedback signal from the listener. Humans are very good in creating opinions about a communication partner based on first impressions from initial meetings. From cross-cultural studies we know that using the wrong social signals in these first encounters easily lead to severe misunderstandings between the communication partners. One aspect of the many social signals is backchannel feedback, specifically head nods. In a previous Japanese study with virtual agents [11] it was shown that the use of head nods helps eliciting more verbal input from the user when they are congruent with culture-specific head nod patterns, in this case for the Japanese culture in contrast to US American patterns. Based on this results, we present a replication of this study here that changes two parameters:

M. Kurosu (Ed.): Human-Computer Interaction, Part II, HCII 2014, LNCS 8511, pp. 651–662, 2014.

(i) The cultural background of the users: Targeting Danish users, we concentrate on Danish nodding patterns based on the analysis of a multimodal corpus of first meeting encounters (NOMCO).
(ii) The embodiment of the agent: Instead of using a virtual character, we replicate the experiment with humanoid robot (Nao), assuming that the physical embodiment will have an impact on the results.

The paper first presents related work in the area of virtual and physical agents. Then the replicated experiment and results are presented. Results show a strong influence of the physical embodiment leading to a follow up study with a virtually present robot, which is presented next before the paper concludes with a discussion.

2 Related Work

Several studies on virtual agents have shown that the paradigm of a listener agent has a good potential of building rapport and engaging the user in prolonged interactions [5–7]. An important aspect is the production and recognition of appropriate social signals in order to realize affective interactions, which are seen as a prerequisite for successfully establishing rapport with the user and it can be safely assumed that this also holds true for interactions with physically embodied agents, i.e. robots. Research on head nods in robots have so far mainly been concerned with recognizing and interpreting head nods by human users (e.g. [8]) but not so much for employing head nods as a means for the robot to structure and maintain the dialogue with the user. Exceptions are the work by [9] and [10].

As has been acknowledged previously, some parameters of head nods seem to vary across cultures like the frequency of head nods in dyadic conversations. Koda and colleagues [11] present an experimental setup for analyzing this cross-cultural variety for a virtual agent system. They showed that human users speak longer to an agent that takes these cultural differences in the realization of head nods into account. Shortcomings of their approach include the fact that they only tested on Japanese subjects. Thus the reported results might be attributable to the fact that more nodding generally elicits more talking from the speaker.

Here we will use the basic experimental setup to test if Danish participants would also prefer to talk longer if a humanoid robot displays culturally adequate feedback signals in terms of head nods. In order to realize this experiment, more information on head nods is necessary. Head movements in general are a well researched feature of human communication focusing on the physical movement itself, on how to classify different movements as well as on the communicative function of the different movements.

McClave [13] distinguishes between two motions for the American culture, an up/down movement (nod) used to signal affirmation and a side to side movement (shake) to signal negation. Allwood and Cerrato [14] present several relevant head movements and distinguish between nod (forward movement of the head going up and down, which can be multiple), jerk (backward movement of the head which is usually single), shake (left-right or right-left movement of the head which can be multiple), waggle (movement of the head back and forth left to right), and 'swturn' (side-way turn is a single turn of the head left or right). Based on these earlier suggestions, Paggio and Navarretta [15] classify head movements into Nod, Jerk, Head-Forward, HeadBackward, Tilt, Side-Turn, Shake, Waggle and HeadOther.

The physical features of head movements have been the focus of Hadar and colleagues [16], who present a number of different results concerning frequency, amplitude, and cyclicity. They report that subjects exhibited head movements every 7.41 seconds on average (frequency of 8.1 movements/minute) without distinguishing between nods, shakes or other movements. The data was also used to analyze the correlation between the amplitude of a head movement and the conversational function, showing e.g. that a mean amplitude of 13.3 degree can be observed with an affirmation ('Yes') and 11.4 degree with a movement that is synchronous to speech. Results are in so far questionable as the means for measuring are very obtrusive and required the subjects to wear a specific head mounted equipment, making the situation far from natural. Also, the analysis is based on a data corpus of around 16 minutes in total. McClave [13] presents some data from Birdwhistell relating to the velocity that can be observed in head nods. She reports the typical velocity range among Americans of 0.8 degrees to 3 degrees per 1/24 second over a spatial arc of 5 to 15 degrees. Maynard [17] is concerned with the frequency and distribution of head nods and presents an in-depth analysis for Japanese dyadic interactions. His analysis reveals that Japanese do one head movement per 5.75 seconds on average (frequency of 10.4 nods/minute) in contrast to Americans with a movement every 22.5 seconds on average (frequency of 2.7 movements/minute). The distribution of head nods between speaker and listener is almost balanced with listeners being responsible for 44% of head nods while speakers are doing 56% of the nods. Paggio and Navarretta [2] present similar data derived from a Danish corpus, which consists of 12 first meeting encounters with a total duration of around 51 minutes. Based on this data, Danish participants nod on average every 5.82 seconds (frequency of 10.3 nods/minute). The above studies reveal a cultural difference in the frequency of head nods, with US Americans nodding less frequently compared to Danish and Japanese.

Apart from information about the physical qualities of head movements, literature on the function of head movements and specifically head nods is vast. In an early study, Dittmann and Lewellyn [12] focus solely on up/down movements, which are recorded by a tailor-made device that the subject had to wear on his head. They attribute two functions to these head nods, either a signal for the speaker that the listener intends to get the floor or as a feedback signal to the speaker. In both cases vocalizations may accompany the head nod, but the head nod may well precede the vocal channel. Heylen [18] gives a comprehensive overview of functions associated with head nods that draws from multiple sources. He distinguishes between 26 different functions but is a bit fuzzy on the use of head nods, as some functions are more associated with gaze than with head nods or include posture changes. In a similar fashion, McClave [13] presents a range of different functions for head nods from semantic over narrative to interactive. Kogure [19] shows that frequent nodding is a phenomenon observed in Japanese conversations whenever a silence in the conversation occurs (so called loop sequence). Thus, they distinguish nods with and without accompanying speech for their analysis. Maynard [17] specifically analyzes Japanese head movements in contrast to American ones and lists the following interactional functions: (1) affirmation; (2) claim for turn-end and turn-transition; (3) pre-turn and turn claim; (4) turn-transition period filler; (5) back channel, and (6) rhythm taking.

Allwood and Cerrato [14] show head movements to be the most frequent feedback signal in dyadic conversations with nodding either single or multiple being by far the most frequent signal they found. In a follow-up analysis Boholm and Allwood [20] show that a majority of multiple head nods accompany speech that also expresses the feedback information (74%).

To sum up, head nods are an important non-verbal feedback signal in human communications. They are found across cultures with variations in their actual realization, e.g. regarding their timing and frequency in an interaction.

3 Online Survey

In order to establish a baseline for the experiment with the humanoid robot, a repeated-measures online survey was conducted to determine which style of head nodding is preferred by Danish users in the context of a listening robot. Head movements were derived from existing video material of students in dyadic first encounter conversations. The videos were analyzed for nodding patterns in velocity, frequency and angle magnitude. Based on these three variables eight varieties of head nods were defined, programmed into a Nao robot and then video recorded. The eight videos shows the robot passively listening to a voice and nodding, where each video depicts a different value combination for the three variables.

After watching each video, participants were asked to report how well they liked the style of head movement according to an 11 point Likert scale. They were also asked to report what emotions they thought the robot seemed to express from a list of 10 arbitrarily selected emotions, equally distributed between positive and negative affect.

Online Survey Results. 41 participants completed the surveys, 24 men and 17 women, with ages ranging from 20 to 57 years (median = 25). The Likert ratings of the videos were analyzed using the Friedman test. The analysis showed no significant results of the rating between the videos. The head nod that was chosen for further experimentation was the head nod that got the highest average Likert score and that correlated with the highest number of positive emotions.

4 Experiment 1: Co-location

From human interaction it is known how feedback positively influences conversation, and the experiment presented in [11] has shown this relates to interactions with a virtual agent. Based on these insights, it can be assumed that users talk longer with a robot which uses culturally appropriate head nods, compared with head nods form another culture or no movement at all. Moreover, it has been shown that speaking activity is a good predictor of the extraversion trait [21]. In order to test these assumptions, an experiment with a physically embodied agent in the form of the Nao robot has been designed with the following hypotheses:

H1. *A robot that nods elicits longer stories from the user compared to one that does not nod.*

H1a. *A robot that shows culture-specific nodding behavior elicits longer stories form the user compared to a robot that shows unspecific or no nodding behavior.*

H2. *Participants scoring high in extraversion will talk considerably longer independent of the experimental conditions compared to user that score low on extraversion.*

H3. *The user will perceive the robot as more intelligent when it elicits backchannel head nods.*

An independent measures Wizard of Oz experiment is conducted. The independent variable in this experiment is the backchannel feedback of the robot. Before the session, participants were informed that they were going to talk to an intelligent robot, that will listen to them but otherwise remain passive. The automatic head-tracking of the Nao was activated to simulate eye contact. Participants were asked to talk to the listing robot about an open-ended, preselected topic from a list of 15 topics[1]. Participants are randomly assigned to either a control group or one of two groups with backchannel feedback. The dependent variable is the duration of how long the participant speaks. Participants are asked to talk to the robot about the chosen topic as long as they can, but for practical purposes are stopped if they speak for more than five minutes. The test leader observes the conversation and heuristically triggers head nods remotely and without the participant knowing. After the test, participants were required to fill out a short questionnaire regarding personality, impression of the robot and demographic. Figure 1 demonstrates the setup.

Participants. Recruiting was done at a 'university college'. All participants were native Danish speaking students with limited knowledge about robots and had various academic backgrounds. They were debriefed after the experiment.

[1] Fifteen topics: fashion, sports, pets, food, books, movies, music, travel, work, studies, games, cars, vacation, hobbies and ambitions

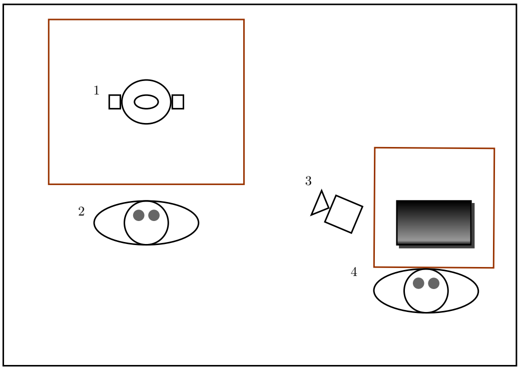

Fig. 1. Top-down sketch of the experiment setup: 1. Nao Robot, 2. Participant, 3. Recording camera, 4. Test facilitator with laptop

Apparatus. The study used a Nao H25 robot by Alderbran robotics. A script was written to trigger the robot to nod upon keyboard input. The remote triggering of the robot was done on a laptop computers with an Intel i7 processor. A video camera was used to record each test session.

Extraversion Measures. Extraversion of each participant was acquired using a shorter version of Eysenck's revisited Eysenck Personality Questionnaire (EPQR-A) [4]. The EPQR-A was administered prior to the each session and only the extraversion dimension was used.

Perceived Intelligence. The participants' perceived intelligence of the robot was obtained using part of Bartneck et al.'s "Godspeed" questionnaire [3]. The questionnaire consists of a series of mutually opposing adjectives, concerning intelligence, working as anchors. The questionnaire consist of five five-point Likert scale questions.

4.1 Results of Co-located Experiment

Twelve female and eight male students participated in the experiment and talked to a physically present robot. Figure 2 shows an example from the test. Their age ranged from 20 to 49, mean = 25, SD = 6.5. Four participants interacted with the robot in the American nodding group and spoke on average 42.8 seconds (SD=9.6). Five participants were in the Danish nodding group and spoke on average 44.6 seconds (SD=9.5). The larger control group (no nodding; NN) had eleven participant who spoke on average 93.8 seconds (SD=28.5). Hypothesis 1 is tested by running two independent measures t-tests between DK-NN and US-NN groups. Bonferroni correction is applied and so $\alpha = (0.05/2) = 0.025$.

The speech duration of participants in the DK group was significantly shorter than the control group $t(14) = 3.7$, $p<0.025$, $r = 0.7$. The speech duration of

Fig. 2. Image of a participant engaging with the robot. The robot remains in the sitting position throughout the experiment and maintains eye contact.

participants in the US group was significantly shorter than the control group $t(13) = 3.4$, $p<0.025$, $r = 0.69$. Contrary to the virtual agent study by Koda et al. the robot elicited less talking from the user when it produced backchannel head nodding. Thus hypothesis 1 is rejected.

To test hypothesis 1a an independent measures t-test is run between DK-US groups. It shows that there is no significant difference between speech duration: $t(7) = 0.289$, $p<0,05$. Thus hypothesis 1a is rejected.

Pearson's correlation coefficient is calculated for the relation between extraversion and duration of speech of a participant in the co-location experiment. There was a positive correlation between the two variables, $r = 0.524$, $n = 20$, $p = 0.018$. A scatter plot of the data is shown if figure 3. Hypothesis 2 is retained.

The results of the Perceived Intelligence are analyzed by comparing the scores of each question between participants of the control group (n=11) and US-DK group combined (n=9). Five independent t-tests, $\alpha = 0.05$, are run. They all showed non-significant difference except for question 5; $t(18) = -2.87$, $p < 0.05$, $r = 0.56$. Thus hypothesis 3 is rejected. While statistically insignificant there was a slight tendency for participants to rate the robot more intelligent on average when they interacted with the robot that elicited feedback. Participants in the nodding-free control condition rated it to be less intelligent.

5 Experiment 2: Virtual Presence

Based on the unexpected outcome of the co-location experiment another experiment is conducted. In this, the independent variable is changed to a virtually present robot that performs Danish head movement to make it more similar to the original study.

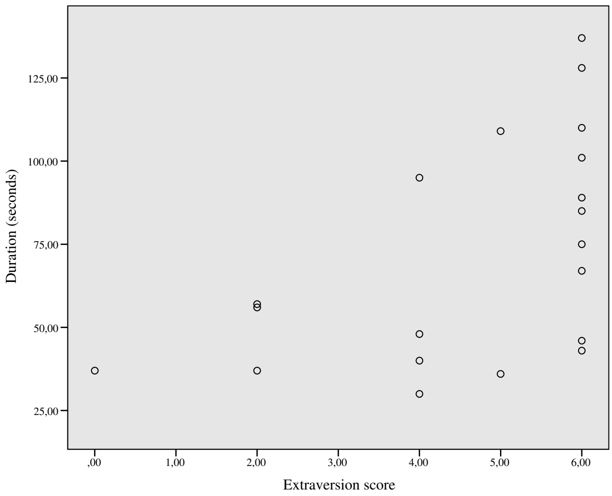

Fig. 3. The measured extraversion of participants in the co-location experiment correlates positively with their duration of speech

The experiment was run with just one condition, the Danish head nodding behavior as a comparison to the previous results. The test was conducted using the same Wizard of Oz method as in the first experiment. The test facilitator is seated with the Nao robot in a separate room and the participants speaks with the robot through a Skype call. The robot performs the same Danish head nod movements as in the first experiment. The same questionnaire data regarding perceived intelligence, extraversion and demography are collected. The following hypothesis is guiding the experimental setup:

H4. *Users will speak considerably longer to a robot that is only virtually present compared to a robot that is physically co-located in the room.*

As in the first experiment participants are asked to speak to a robot about one of the 15 topics. They are given the same instructions as in the first experiment except they are required to answer a Skype call with the robot. The test conductor leaves the room with the explanation that he is monitoring the Skype call remotely.

5.1 Participants

Nine participants could be won, 4 females and 5 males, all native Danish speakers with an age range from 22 to 64 years (median = 25) of which the majority were students.

5.2 Apparatus

The participant laptop had a 15.6 inch screen and was placed on a table in front of the participant. An external microphone is plugged into the laptop and placed

Fig. 4. Image of a participant engaged in a Skype conversation with the robot. The robot remains in the same sitting position as before but only upper body and head is visible on the screen.

in front of the laptop to ensure the facilitator clearly hears the participant during the test. Figure 4 shows the setup of the Skype experiment.

5.3 Results

The 9 participants in the virtual presence experiment spoke on average for 204,7 seconds (SD = 93.2). This is compared in an independent measures t-test, $\alpha = 0.05$, with the results of participants in the co-location experiment in the DK group (n = 5, mean = 44.6, SD = 9.5). Participants in the virtual presence experiment spoke significantly longer: $t(8.29) = -5.08$, $p < 0.05$. The hypothesis H4 is thus retained. Participants' average extraversion was 5.5.

5.4 Discussion

Participants indeed talked significantly longer when the robot was not physically present in the room, on average more than twice as long as in the control condition. This is also in line with the findings by Koda and colleagues for the virtual agent case [11]. It should be noted that participants in this group had high extraversion which may partially account for the high duration. Thus, we can conclude the physical presence of the robot has a huge impact on the users' behavior, at least in cases where users meet a robot for the first time in their life. It remains to be shown if this effect vanishes, when users get more familiar with the robot, e.g. in subsequent sessions.

6 Discussion and Limitations

The results of the first study contradict the assumptions made from human communication and previous studies with virtual agents. The head movement of the robot negatively influenced the speech duration of participants compared to participants who spoke to the robot that did not produce any head feedback. We speculate that the physical presence of a robot is the cause of this outcome, as a contrast to the virtual agent of the original Japanese study. The difference between co-located and virtual presence, could have been influenced by the presence of the test facilitator during the co-located test and have caused discomfort of the participants. On the other hand, this would not explain why participants spoke longer in the control condition (no nod). The experiment noticeably differs in that it uses a robot, compared to the original experiment which uses a virtual character. We assume that the results might be attributed to either the use of a robot or the presence of the test facilitator during the test, encouraging for future experiments.

7 Conclusion

Danish participants spoke to a robot under different condition of presence and backchannel feedback. Contrary to our hypothesis the duration of speech was significantly shorter when the robot produced head movement compared to a control condition with no head nods. There was no significant difference in speech duration whether the backchannel feedback of the robot was culture specific. As expected a positive correlation was found between speech duration and extraversion. Participants spoke significantly longer when the robot was virtually present. It is not trivial to replicate human communication in interaction with robots. Nor is it a matter of simply reproducing communication signals in a robot to make users interact with it as if it were human.

Acknowledgements. The present work benefitted from the work of Constanza Navaretta and Patrizia Paggio researchers at "Center for Sprogteknologi" at Copenhagen University who provided the NOMCO data. Moreover the experiments benefits from the help of Vagn E. Pedersen, Sarah Catterine, Birthe Brøndum and Harry Brøndum for their help in running the experiments described in this paper.

References

1. Paggio, P., Navaretta, C.: Feedback and gestural behaviour in a conversational corpus of Danish. In: Proceedings of the 3rd Nordic Symposium on Multimodal Communication NEALT (2011), pp. 33–39 (2011)
2. Paggio, P., Navarretta, C.: Head Movements, Facial Expressions and Feedback in Danish First Encounters Interactions: A Culture-Specific Analysis. In: Stephanidis, C. (ed.) Universal Access in HCI, Part II, HCII 2011. LNCS, vol. 6766, pp. 583–590. Springer, Heidelberg (2011)

3. Bartneck, C., Kuli, D., Croft, E., Zoghbi, S.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics 1, 71–81 (2009)
4. Francis, L., Brown, J., Philipchalk, L.B., The, R.: The development of an abbreviated form of the revised eysenck personality questionnaire (EPQR-A): Its use among students in England, Canada, the U.S.A. and Australia. Personality and Individual Differences 13(4), 443–449 (1992)
5. Gunes, H., Heylen, D., ter Maat, M., McKeown, G., Pammi, S., Pantic, M., Pelachaud, C., Schuller, B., de Sevin, E., Valstar, M., Wöllmer, M.: Building Autonomous Sensitive Artificial Listeners. IEEE Transactions on Affecite Computing 3, 165–183 (2012)
6. Yoichi, S., Yuuko, N., Kiyoshi, Y., Yukiko, N.: Listener Agent for Elderly People with Dementia. In: Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, pp. 199–200. ACM, New York (2012)
7. Meguro, T., Higashinaka, R., Dohsaka, K., Minami, Y., Isozaki, H.: Analysis of listening-oriented dialogue for building listening agents. In: Proceedings of the SIGDIAL 2009 Conference, pp. 124–127. Association for Computational Linguistics, Stroudsburg (2009)
8. Rich, C., Ponsler, B., Holroyd, A., Sidner, C.: Recognizing engagement in human-robot interaction. In: Proceedings of HRI, pp. 375–382. Institute of Electrical and Electronics Engineers (IEEE) (2010)
9. Riek, L.D., Paul, P.C., Robinson, P.: When my robot smiles at me Enabling human-robot rapport via realtime head gesture mimicry. Journal of Multimodal User Interfaces, vol 3, 99–108 (2010)
10. Liu, C., Ishi, C., Ishiguro, H., Hagita, N.: Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction. In: Human-Robot Interaction, pp. 285–292. ACM (2012)
11. Koda, T., Kishi, H., Hamamoto, T., Suzuki, Y.: Cultural Study on Speech Duration and Perception of Virtual Agent's Nodding. In: Nakano, Y., Neff, M., Paiva, A., Walker, M. (eds.) IVA 2012. LNCS (LNAI), vol. 7502, pp. 404–411. Springer, Heidelberg (2012)
12. Dittmann, A., Llewellyn, L.: Relationship between vocalizations and head nods as listener responses. Journal of Personality and Social Psychology 9, 79–84 (1968)
13. McClave, E.: Linguistic functions of head movements in the context of speech. Journal of Pragmatics 32, 855–878 (2000)
14. Allwood, J., Cerrato, L.: A study of gestural feedback expressions. In: Paggio, P., Jokinen, K., Jönsso, A. (eds.) First Nordic Symposium on Multimodal Communication, Copenhagen, pp. 7–22 (2003)
15. Paggio, P., Navarretta, C.: Head movements, facial expressions and feedback in conversations: Empirical evidence from Danish multimodal data. Journal on Multimodal User Interfaces 7, 29–37 (2013)
16. Hadar, U., Steiner, T., Rose, F.: Head movement during listening turns in conversation. Journal of Nonverbal Behavior 9, 214–228 (1985)
17. Maynard, S.: Interactional functions of a nonverbal sign Head movement in japanese dyadic casual conversation. Journal of Pragmatics 11, 589–606 (1987)
18. Heylen, D.: Challenges ahead: head movements and other social acts during conversations. In: Joint Symposium on Virtual Social Agents. AISB (2005)

19. Kogure, M.: Nodding and smiling in silence during the loop sequence of backchannels in Japanese conversation. Journal of Pragmatics 39, 1275–1289 (2007)
20. Boholm, M., Allwood, J.: Repeated head movements, their function and relation to speech. In: Kipp, M., Martin, J., Paggio, P., Heylen, D. (eds.) Proceedings of the Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality, pp. 6–10. LREC (2010)
21. Lepri, B., Subramanian, R., Kalimeri, K., Staiano, J., Pianesi, F., Sebe, N.: Employing social gaze and speaking activity for automatic determination of the extraversion trait. In: International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, p. 7. ACM (2010)