Perception of a humanoid robot as an interface for auditory testing

Luke Meyer[1,2] [0000-0001-7340-3802], Gloria Araiza-Illan[1,2] [0000-0001-7539-5324], Laura Rachman[1,2][0000-0003-4744-1308], Etienne Gaudrain[1,3] [0000-0003-0490-0295] and Deniz Başkent[1,2] [0000-0002-6560-1451]

[1] Department of Otorhinolaryngology / Head and Neck Surgery, University Medical Centre Groningen, University of Groningen, The Netherlands
[2] University of Groningen, University Medical Centre Groningen, W.J. Kolff Institute for Biomedical Engineering and Materials Science, Groningen, the Netherlands.
[3] Lyon Neuroscience Research Centre, CNRS UMR 5292, INSERM UMRS 1028, Université Claude Bernard Lyon 1, Université de Lyon, Lyon, France

Abstract. Perception tasks that require long and frequent testing can result in a loss of attention and focus, contributing to erroneous and inconsistent data. To maintain engagement and improve reliability of data, a more interactive interface could be used. This study aims to investigate if a humanoid NAO robot could provide such an interface to improve or maintain engagement during testing. More specifically, human-robot interaction (HRI) will be explored when performing various auditory perception tasks, played in the form of games, on the NAO robot. Evaluation of the HRI will be performed using questionnaires derived from the system usability, Godspeed, and similarity-attraction scales, as well as through video analysis. Future research will extend the evaluation of the HRI to children and aim to further improve the engagement and social acceptability of the NAO robot as a game interface.

Keywords: HRI, auditory perception, robot perception, task engagement.

1 Introduction

Maintaining focus and engagement during perception tasks that require long or frequent repetitions can become challenging for individuals with limited attention spans, such as children [1]. More interactive interfaces, such as humanoid robots, could help, as they have been shown to be adept at retaining attention in comparison to more commonly used laptops [2]. As a preliminary investigation into engagement when using a humanoid robot, this study aims to explore the impression adults have of a NAO V5 humanoid robot as an interactive interface for games designed for auditory perception testing. The Perception of Indexical Cues in Kids and Adults (PICKA) test battery [3] is a set of auditory perception tasks, on perception of voice and speech, as an ongoing investigation into hearing loss. To obtain reliable data, these tests – played in the form of games on a laptop – are performed frequently over

relatively long testing times, often resulting in a loss of concentration; thus, necessitating a more engaging interface. Maintaining engagement during the PICKA games could lead to longer testing times, providing more consistent data.

2 Experimental Design

Thirty normal hearing, proficient English speaking, locally recruited adult participants, both from Dutch and international backgrounds, thus far between the ages of 19–38 (24.04±5.26) years, play one of the four PICKA games on either the laptop or NAO interface, randomly determined, followed by the same game on the other interface. Due to the game duration and repetitive procedure, the study is divided into two experiments in which two of the games are paired together. Which games participants play is also randomly determined; however, counterbalancing was performed across both experiments for games and interfaces. When using the NAO, it both plays the stimuli for the game and logs responses, given through the tactile sensors, from participants. In comparison, when using the laptop, stimuli are presented from the laptop speakers and responses are given using the laptop screen and external mouse.

Each of the four PICKA games is composed of an introduction, a training phase, a testing phase, and breaks (where applicable). At the start of each game, NAO stands up, introduces the game, and familiarizes the participant with how and where to touch the tactile sensors. Following the introduction, NAO returns to a seated position for the rest of the game. In this position, the motors and fans are switched off, making NAO much quieter and the auditory stimuli easier to hear.

Prior to the experiment, participants are asked to complete the revised personality index (neo-PIR) questionnaire [4] to obtain an indication of their extroversion/introversion, as this would provide an additional parameter for analysing how one's personality influences their interaction with the robot. If a trend exists, this could potentially be used to adapt the interaction based on an individual's personality [5] to improve the level of engagement, as well as to enhance the interaction. Additionally, the Negative Attitude towards Robots Scale (NARS) [4] questionnaire is included to account for potential factors that may influence the interaction and used in covariate analyses. Video recordings of the participants are taken from two cameras of both the laptop and NAO versions of the games. One video camera is placed behind the laptop/NAO to capture the face of the participant, and another to the side of the participant to capture the participant's movements and interactions with the interfaces. After the completion of the game on both interfaces, the participant is asked to complete a new set of questionnaires to evaluate their experience with the robot and the laptop. Questionnaires include the system usability scale (SUS) [6], questions about the experiment, the Godspeed questionnaire [7], and the similarity-attraction questionnaire [8]. Evaluation of the HRI is performed using both questionnaires and video recordings.

2.1 Voice cue sensitivity

Three pseudowords are presented, one of which sounds different to the other two in a three-interval three-alternative-forced-choice paradigm. Participants must identify which of the three stimuli was different by touching NAO's right hand, head, or left hand (for first, second and third stimuli, respectively). The acoustic difference between the stimuli becomes progressively smaller as the participant answers correctly. If an incorrect response is given, the acoustic difference becomes greater, making it easier to discern. To obtain reliable data, this paradigm is presented four times separated by a short break, totalling 30–40 minutes. When NAO is used, it offers a break to the participant, to which they verbally reply. If they take the break, NAO offers them to join in a stretch routine. If not, NAO remains seated for a short time before asking them if they are ready to continue. In the laptop version of the game, progress is displayed with a progress bar, as well as a running tally of correct responses. The displaying of a progress bar is not currently implemented on NAO, and thus to accommodate for the lack of progress tracking, NAO praises participants if they provide consistent correct answers or motivates them if incorrect responses are given. NAO also provides visual feedback for each answer, nodding for a correct response or shaking its head for an incorrect response.

2.2 Gender categorization

An English word is presented to the participant, and they must subjectively categorize the gender (male or female) of the spoken voice. After the stimulus is presented, NAO indicates which of its hands can be touched for which gender (they are randomized after each stimulus to avoid association of a gender to a specific hand). No visual feedback is presented to the participant after their responses. The eyes, however, do change colour to indicate when a response can be given, and again when the response has been stored. Since this game is much shorter (8–10 minutes) than the voice cue sensitivity, no breaks are offered.

2.3 Emotion identification

This game presents the participant with a nonsensical sentence spoken with either a happy, angry, or sad voice. The participant uses NAO's hands and head to input their responses. In comparison to the gender categorization game, the hand-emotion pair is kept constant as it could confuse participants if randomized each time, or significantly increase the duration if NAO indicates the pair after each stimulus. Identifying the emotion of the voice is not subjective; thus, visual feedback is provided after an answer is given. Again, no breaks are offered to the participant as the game is relatively short (5–10 minutes).

2.4 Speech-on-speech perception

This game uses an adapted version of the coordinate response measure (CRM) [9]. A sentence containing a colour and number is presented to the participant, who uses a tablet with a coloured and numbered grid to indicate the heard colour and number. A tablet is used here as there are more combinations of colours and numbers than inputs on NAO. The stimuli also contain masker signals to simulate background speech, varying the difficulty of identifying the colour and number. Half-way through the game (total duration is 15–20 minutes), NAO offers an optional break to the participant, identical to that described above. Visual feedback is provided to the participant based on their responses.

2.5 Video analysis

A combination of social cues described by Giuliani et al. [10] and Desideri et al. [11], including both verbal and non-verbal cues, are used to code videos of both the laptop and robot interfaces. Segments between one to two minutes from each part of the PICKA games are taken for each participant, which are then randomized and merged into a single video. Four reviewers are used to code the videos, which are divided such that each game from each participant is coded by at least two reviewers in a fully crossed coding design.

3 Further work

As this study is still underway, results cannot yet be reported. However, it is expected that the perception of NAO as the interface for the PICKA games will be favoured over the laptop interface. It is also expected that the results of the questionnaires will correlate with the video analysis of the interaction; i.e., a higher useability score, similarity-attraction, and likeability of NAO will also be present as longer maintained engagement and improved enjoyment with the robot. Furthermore, it is expected that the results of the neo-PIR would give an indication toward future adaptations; i.e., how the interaction could be improved for introverted/extroverted individuals.

Further investigation will include playing the PICKA games on the NAO robot with more vulnerable populations, starting with normal hearing children followed by hard-of-hearing children. This will provide additional data on how the interaction can be further adapted to better establish and maintain engagement with them. Anticipated challenges regarding the next phase primarily concern children becoming distracted when using the robot and thus contradicting the intention of the NAO robot, or the duration of the test protocol being too long for the limited attention spans of children.

References

1. Gaudrain E, Başkent D. Discrimination of Voice Pitch and Vocal-Tract Length in Cochlear Implant Users: Ear Hear. 2018;39(2):226–37.

2. Looije R, van der Zalm A, Neerincx MA, Beun R-J. Help, I need some body the effect of embodiment on playful learning. In: 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication [Internet]. Paris, France: IEEE; 2012 [cited 2021 Feb 15]. p. 718–24. Available from: http://ieeexplore.ieee.org/document/6343836/
3. Nagels L, Gaudrain E, Vickers D, Hendriks P, Başkent D. Development of voice perception is dissociated across gender cues in school-age children. Sci Rep. 2020 Dec;10(1):5074.
4. Ivaldi S. Towards Engagement Models that Consider Individual Factors in HRI: On the Relation of Extroversion and Negative Attitude Towards Robots to Gaze and Speech During a Human–Robot Assembly Task. 2017;24.
5. Bernier EP, Scassellati B. The similarity-attraction effect in human-robot interaction. In: 2010 IEEE 9th International Conference on Development and Learning [Internet]. Ann Arbor, MI, USA: IEEE; 2010 [cited 2021 Feb 26]. p. 286–90. Available from: http://ieeexplore.ieee.org/document/5578828/
6. Brooke J. SUS - A quick and dirty usability scale. In: Jordan PW, Thomas B, Weermeester BA, McClellend AL, editors. Usability evaluation in industry. London: Taylor and Francis; 1996.
7. Bartneck C, Kulić D, Croft E, Zoghbi S. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. Int J Soc Robot. 2009 Jan;1(1):71–81.
8. Lee KM, Peng W, Jin S-A, Yan C. Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction. J Commun. 2006 Dec 1;56(4):754–72.
9. Eddins DA, Liu C. Psychometric properties of the coordinate response measure corpus with various types of background interference. J Acoust Soc Am. 2012 Feb;131(2):EL177–83.
10. Giuliani M, Mirnig N, Stollnberger G, Stadler S, Buchner R, Tscheligi M. Systematic analysis of video data from different human–robot interaction studies: a categorization of social signals during error situations. Front Psychol [Internet]. 2015 Jul 8 [cited 2021 Jan 4];6. Available from: http://journal.frontiersin.org/Article/10.3389/fpsyg.2015.00931/abstract
11. Desideri L, Ottaviani C, Malavasi M, di Marzio R, Bonifacci P. Emotional processes in human-robot interaction during brief cognitive testing. Comput Hum Behav. 2019 Jan;90:331–42.