More Human than Human? A Visual Processing Approach to Exploring Believability of Android Faces

Masayuki Nakane
University of Manitoba
umnakane@myumanitoba.ca

James E. Young
University of Manitoba
young@cs.umanitoba.ca

Neil D. B. Bruce
University of Manitoba
bruce@cs.umanitoba.ca

ABSTRACT

The issue of believability is core to android science, the challenge of creating a robot that can pass as a near human. While researchers are making great strides in improving the quality of androids and their likeness to people, it is simultaneously important to develop theoretical foundations behind believability, and experimental methods for exploring believability. In this paper, we explore a visual processing approach to investigating the believability of android faces, and present results from a study comparing current-generation android faces to humans. We show how android faces are still not quite as believable as humans, and provide some mechanisms that may be used to investigate and compare believability in future projects.

Author Keywords

Android Science; Human-Robot Interaction; Visual Processing; Uncanny Valley

ACM Classification Keywords

H.1.2 Models and Principles: User/Machine Systems. Software Psychology

INTRODUCTION

Androids are a class of robots that have the ultimate goal of being able to approximately pass for human [16]. To accomplish this, androids will eventually have to look, move, and interact the same as everyday people. When these goals are not met, not only are androids easily identifiable as non-human, but interaction can suffer in other ways, such as the android appearing eerie or making people uncomfortable (often referred to the Uncanny Valley problem [15]). Moving forward, android developers will need a solid understanding of which features and characteristics impact believability of their androids, understanding of the underlying perception mechanisms that impact believability, and tools and methods to help diagnose and determine their own android's believability.

HAI '14, October 28 - 31 2014, Tsukuba, Japan

ACM 978-1-4503-3035-0/14/10…$15.00.
http://dx.doi.org/10.1145/2658861.2658893

In this paper, we take an initial step toward this goal by exploring visual perception and processing of android faces. We purposely select a heavy simplification of the broader problem, focusing on the visual perception of static images of android faces. This serves as an initial base case where believability is arguably more easily achievable than with real robots, motion, and interaction.

We present a visual processing discussion and initial foundations explaining how people may process android faces, and conducted a study based on this theory comparing human faces to faces of current-generation androids. While androids are becoming impressively believable, our results show that – as predicted by our theory – people are still faster at identifying human faces, find android faces more eerie, and make more mistakes with android faces.

This work provides an initial step toward building a theoretical foundation for the believability of android faces. At the very least, we have shown how simple studies examining face-identification times and error rates can be used to test android faces and infer potential believability.

RELATED WORK

The general study of how robotic design impacts interaction is well established in the field, e.g., comparing zoomorphic and anthropomorphic designs in terms of perceived animacy [1], or building frameworks for appropriate and believable social robot behaviors (e.g., [17]). We propose to extend this direction by specifically addressing the believability of android faces.

Most work on the believability of robots surrounds the eeriness problem (often called the uncanny valley problem). Since first postulated [22], this issue has been contentious [13], and many researchers have looked to unpack the issue in terms of robot dimensions, e.g., morphology [2, 8, 15], or realism and iconicity [5, 6]. While eeriness is inevitably a part of believability of android faces, we take a more holistic approach where eeriness is but one part of the issue.

There has been limited work in robotics and animation that looks at how people visually process artificial faces. One work looked at how people process a real face, an animated face, and various points "morphed" in between [8], and found that people took longer to classify ambiguous faces than clearly human or animated, with classification time decreasing as the ambiguity decreased. In our work we aim to continue this direction and explicitly target androids.

Some researchers have looked at the impact of faces, for example how modifying facial features can impact response such as perceived attractiveness of an agent [9], and similar work looked at how people can apply gender stereotypes to a robot based solely on the haircut [12]. We extend this direction by addressing believability.

VISUAL PROCESSING

Humans are hard wired to see and find faces, even where none exist: this can be illustrated by a common face-finding exercise in doodles [19], and has been linked to a human need for social interaction [10] – people can recognize a face in about 350ms [24]. As such, one would expect the creation of believable android faces to be fairly simple. Unfortunately, although people can easily understand even a crude android design as having a face, believability is a separate problem.

When a person sees a human face – or a face they believe may be human, such as an android – they conduct face-recognition processing, to determine if they know the face. Unlike the simple task of finding faces, this requires a great deal more sophistication given the range of differences and subtleties between people's faces. There is a body of work examining this processing, much of it dealing with how people scan and fixate on a potential face (e.g., [3, 24]).

When a person detects a face which at first-glance appears to be human, but deeper face-recognition processing detects a problem, we have an "expectancy violation" [21]. Such violations in processing draws a person's attention (even at the subconscious level) to the violation as a means of investigating why the violations happened; as such, we can expect that faces which are not quite normal will take more time for a person to process, given this violation. Further, expectancy violations have also been shown to impact anthropomorphism and believability in other contexts [21], and so we can anticipate similar results here. Arguably, if this violation is jarring and the drop in anthropomorphism is large, this may contribute to the eeriness problem.

There are standard ways that people scan faces, for example, many people initially look below the eyes [24], and usually look at the nose, mouth, and cheeks in some order [3], commonly forming a distinct "T" pattern (eyes and then down) [23]. There is also evidence that problems with eyes are more salient than other features [14]. Individual differences (e.g., based on culture [18] or gender [4]) do exist, which is important to consider for studying believability of androids; for example, some studies show women as having superior face processing [4] so they may be more difficult to deceive with androids. People also have varying tendencies to anthropomorphize (dispositional anthropomorphism) – to give non-human things such as images and potential faces human qualities [11]. Despite these differences, it may still be feasible to study general eye-gaze patterns to help diagnose why an android face is not seen as human. For example by detecting uncharacteristically long fixations or fixation order across many people. In addition, it may be useful to measure a person's disposition to anthropomorphize as an important source of error in data analysis, where people with lower disposition may perhaps be better at detecting issues with an android face.

Some research has purposely distorted human faces to study results and infer about visual processing. For example, by inverting faces (upside-down) or components (inverting the mouth or ryes only), to separate whole-face from component processing [7]. Such techniques, including as hiding the eyes or mouth respectively and doing recognition tests (e.g., as in [7, 14, 25]), can be useful to diagnose components of an android face.

EXPLORATORY PILOT STUDY

We conducted an initial study looking at people's processing of android and human faces. As a pilot, we focused simply on people's classification of static images of faces as either android or human, following the experiment design of [8], and conducted a series of exploratory analyses.

As a primary base case, we wanted to investigate if current-generation android faces are sufficiently believable as human. Also, based on our visual processing background given above we hypothesized that people would take longer to process and classify android faces, would make more mistakes (higher error rate), and, due to the increased ambiguity, would find the android faces more eerie. We anticipated that a person's disposition toward anthropomorphism (general tendency to anthropomorphize) would negatively correlate with response time, as we postulated that they would more readily accept the android face as human, and would have higher error rates with android faces given their potential tendency to mistake them for human. Further, we expected that female participants would have lower error rates and quicker response times due to potential face recognition advantages.

Methodology and Procedure

We recruited fourteen participants from our general university population aged 18-58 (Mdn=20.5), with an equal male / female split, and paid them $10 for their participation. Participant nationalities were primarily Canadian, and also included Nigerian, Chinese, Brazilian, and Pakistani participants.

Experiments were conducted with one participant at a time. Participants were briefed about the study and we obtained informed consent. They then sat at a desk at a fixed distance (15cm) from a 24" wide-screen 16:10 monitor and fixed location (centered); we used a chin-rest (sanitized between participants) to ensure this, and participants wore a light-weight eye-tracking device (PT mini) – this was for technical pilot reasons only and the eye-tracking data was not used in our study. Participants completed the tasks where they classified faces as either human or android, and finished with a post-test questionnaire. The entire study took roughly 30 minutes.

Tasks

Our experiment consisted of two tasks: 1) classification accuracy priority, and 2) classification speed priority. In task 1, participants were shown a face for three full seconds, after which they were asked to classify it and were verbally asked post-stimulus questions before being shown the next face. The order of faces was counterbalanced using an incomplete Latin square. This design enabled the person to concentrate on the face and not to feel rushed in their decision making. In task 2, participants were shown the same faces (with a different counterbalanced order), but were asked to classify them as quickly as possible while the face was shown, with the response time being digitally recorded. The rest of the presentation style was the same except there were no post-stimulus questions here.

In both cases, faces were shown at random locations on the screen, and a blank screen with a fixation cross in the center was placed between faces (during questions) to minimize cross-over effect. Further, in both cases the participant held a mouse in their hand with thumbs on the two buttons, and use these to classify the faces by pressing one of them (left for human, right for android).

Instruments

We compiled a database of faces, consisting of ten android and ten human faces, with half of each category being female. Faces were selected as much as possible to have a neutral expression and to be fully front-facing. Figure 1 shows faces used and provides source attributions. All images were scaled to 324x386 for consistency across faces, which was 3.4" x 4.5" on our screen. For the post stimulus verbal questions, for task 1, we asked the participant to rate how "eerie" the image was on a scale of 1-7.

The post-test questionnaire collected basic demographics, and included the Individual Difference in Anthropomorphism Questionnaire [24] to measure the participant's disposition toward anthropomorphism.

Results

Based on task 1 results, overall androids were more likely to pass for human (57%) than android (43%), and humans were uncommonly mistaken for androids (10%), $\chi^2(1)=38.87$, $p<.001$, and $\chi^2(1)=2.86$, $p<.1$ for android faces only (Table 1). Despite this result, however, there were marked differences in how the faces were processed.

We used one-tailed, paired t-tests to further compare the results on android and human faces. On average, participants classified human faces faster ($M=1.66$s, $SE=.14$) than android faces ($M=2.00$s, $SE=.14$), ($t_{13}=-4.16$, $p<.01$, $d=.7$), reported that human faces were less eerie ($M=2.21$, $SE=.33$) than android faces ($M=2.57$, $SE=.36$, $t_{13}=1.87$, $p<0.05$, $d=.28$), and participants were found to have a lower error rate for classifying humans ($M=.12$, $SE=.06$) than androids ($M=.51$, $SE=.07$), ($t_{13}=3.77$, $p<.01$, $d=1.65$).

We performed correlation tests between the IDAQ (disposition toward anthropomorphism) questionnaire answers and other results, but all results were non-significant ($p>.50$). In addition, no effect of participant gender or face gender was found on any measure (F<1).

Discussion

Our results show that people classified android faces as human faces at the confidence level of 90% (or $\alpha=.1$). At

Table 1 – cross tabulation of how human and android faces were classified, $\chi^2(1)=38.87$, p<.001.

Count

		classified as		
		human	android	Total
face	human	126	14	140
	android	80	60	140
Total		206	74	280

Figure 1 – Face images used in our study, human faces on the left and android faces on the right. The human faces are extracted from the FEI face database (http://fei.edu.br/~cet/facedatabase.html). We compiled the android faces through sources available online. From left to right, top row first: Hanson Robotics' Philip K. Dick, Bina 48, Jules (http://www.hansonrobotics.com/), ATR Geminoid and Geminoid F. Second row, FaceTeam FACE robot [18], JST ERATO Asada and Kokoro CB², Neurobiotics Alissa, KITECH Ever-2, National Taiwan University Robot (unnamed).

least for the simplified problem space of static images, we believe this demonstrates that android faces are doing quite well in terms of believability. Further, as expected, participants classified human faces faster than android faces, had a lower error rate, and found them less eerie. This shows that, even when android faces may pass for human, there are elements of visual processing and response that can highlight the differences between human and android faces. This initial pilot result lends support to our visual processing approach, and highlights how it can be applied to gain insight into believability of faces. For example, that through "expectancy violation" faces that have issues will take longer to process, and will be more ambiguous.

As the android faces were seen as being more eerie than the human faces, post-hoc we performed correlation tests between eeriness and the error rate and response time, to see if eeriness may predict the other factors. Unfortunately these tests were not significant. We believe that it will be important to continue to investigate how eeriness relates to visual processing and believability of faces.

The lack of results relating to disposition to anthropomorphize was surprising. Given the very poor results (t<1 in most cases, illustrated in Figure 2), we do not feel that this would become significant with more participants with our current setup. However, it is difficult to determine if the effect did not exist, if our sample size was too small, if other factors were larger than tendency to anthropomorphize, or if our IDAQ questionnaire did not measure it well, and so we encourage further inquiry in this area. Similarly, we suggest further inquiry into the effect of gender (both the person's [4] and the android's) on face believability.

LIMITATIONS

A key limitation of this study was the small sample size. Further work in the area must address this to find results that are more generalizable.

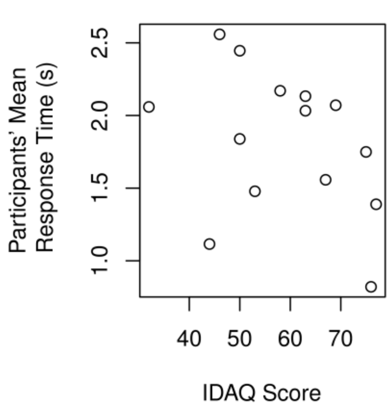

Figure 2 – Participant mean response time (s) against tendency to anthropomorphize (IDAQ score). As shown, there was little relationship found.

The primary purpose of this paper is to explore the visual processing approach to considering the believability of android faces, and so our primary limitation is the small scope of the work. We hope that this direction continues to be developed, and stronger android-centric visual processing theories can be developed to better inform design. From this, we hope that further studies will continue to be conducted to unpack the complexity of android face believability. For example, one element of our theory which we did not address yet is the gaze element of facial processing – how people fixate and process a face.

One facet which must be addressed for continuing work in this area is the development of a standardized face database. Our results are deducted from only ten faces of android faces. While we attempted to maintain uniform lighting, angle, and size across them, taking more care to develop such a database with more faces would provide an excellent benchmark for researchers to compare against. Such a database could include meta-data such as the race, supposed age, and gender of the face.

CONCLUSION

Understanding why a particular android's face is believable, and what can be done about it, is an important challenge for android science. While this is a large challenge, in this work we provided a new angle on the problem, explicitly looking to visual processing knowledge to understand how people are viewing faces. Using this, we have discovered how recognition time and error rate, as well as perhaps perception of eeriness, can all be indicators of believability of an android face as human. In addition, through our exploration we have highlighted various other future directions for explorations in this area.

REFERENCES

[1] Bartneck, C. et al. 2009. Does the Design of a Robot Influence Its Animacy and Perceived Intelligence? *International Journal of Social Robotics*. 1, 2 (Feb. 2009), 195–204.

[2] Bartneck, C. et al. 2007. Is the uncanny valley an uncanny cliff? *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication* (2007), 368–373.

[3] Barton, J.J.S. et al. 2006. Information processing during face recognition: the effects of familiarity, inversion, and morphing on scanning fixations. *Perception*. 35, (2006), 1089–1105.

[4] Bate, S. et al. 2010. Socio-emotional functioning and face recognition ability in the normal population. *Personality and Individual Differences*. 48, 2 (2010), 239–242.

[5] Blow, M., Dautenhahn, K., Appleby, A., Nehaniv, C. L., & Lee, D. (2006, March). The art of designing robot faces: Dimensions for human-robot interaction. In Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction (pp. 331-332). ACM.

[6] Blow, M., Dautenhahn, K., Appleby, A., Nehaniv, C. L., & Lee, D. C. (2006, September). Perception of robot smiles and dimensions for human-robot interaction design. In Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on (pp. 469-474). IEEE.

[7] Bombari, D. et al. 2009. Featural, configural, and holistic face-processing strategies evoke different scan patterns. *Perception*. 38, (2009), 1508–1521.

[8] Cheetham, M. et al. 2013. Category Processing and the human likeness dimension of the Uncanny Valley Hypothesis: Eye-Tracking Data. *Frontiers in psychology*. 4, March (2013), 108.

[9] Chen, H. et al. 2010. Crossing the “uncanny valley”: adaptation to cartoon faces can influence perception of human faces. *Perception*. 39, 3 (2010), 378–386.

[10] Epley, N. et al. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychological review*. 114, (2007), 864–886.

[11] Epley, N. et al. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychological review*. 114, (2007), 864–886.

[12] Eyssel, F. and Hegel, F. 2012. (S)he’s Got the Look: Gender Stereotyping of Robots1. *Journal of Applied Social Psychology*. 42, 9 (Sep. 2012), 2213–2230.

[13] Hanson, D. et al. 2005. Upending the Uncanny Valley. *Proceedings of the national conference on artificial intelligence*. 20, 4 (2005), 24–31.

[14] Hills, P.J. et al. 2013. First fixations in face processing: The more diagnostic they are the smaller the face-inversion effect. *Acta Psychologica*. 142, (2013), 211–219.

[15] Ho, C. et al. 2008. Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. *... conference on Human robot ...*. (2008), 169–176.

[16] Ishiguro, H. 2007. Android Science - Toward a new cross-interdisciplinary framework -. *Robotics Research: Results of the 12th International Symposium ISRR. Springer, 2007* (2007), 118–127.

[17] Kitade, T. et al. 2013. Understanding suitable locations for waiting. *ACM/IEEE International Conference on Human-Robot Interaction* (2013), 57–64.

[18] Mazzei, D. et al. 2012. HEFES: An Hybrid Engine for Facial Expressions Synthesis to control human-like androids and avatars. *Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics* (2012), 195–200.

[19] McCloud, S. 1994. *Understanding Comics*.

[20] Miellet, S. et al. 2010. Investigating cultural diversity for extrafoveal information use in visual scenes. *Journal of vision*. 10, (2010), 21.

[21] Morewedge, C.K. 2009. Negativity bias in attribution of external agency. *Journal of experimental psychology. General*. 138, (2009), 535–545.

[22] Mori, M. et al. 2012. The uncanny valley. *IEEE Robotics and Automation Magazine*. 19, 2 (2012), 98–100.

[23] Peterson, M.F. and Eckstein, M.P. 2013. Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological science*. 24, (2013), 1216–25.

[24] Peterson, M.F. and Eckstein, M.P. 2012. Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences of the United States of America*. 109, (2012), E3314–23.

[25] Schwaninger, A. et al. 2006. Chapter 18 Processing of facial identity and expression: a psychophysical, physiological, and computational perspective. *Progress in Brain Research*.

[26] Waytz, A. et al. 2010. Who Sees Human? *Perspectives on Psychological Science*. 5, (2010), 219–232.