Android Emotions Revealed

Evgenios Vlachos and Henrik Schärfe

Aalborg University, Department of Communication and Psychology,
Nyhavnsgade 14, DK-9000
{evlach10@student , scharfe@hum}.aau.dk

Abstract. This work presents a method for designing facial interfaces for sociable android robots in respect to the fundamental rules of human affect expression. Extending the work of Paul Ekman towards a robotic direction, we follow the judgment-based approach for measuring facial expressions to test in which case an android robot such as the Geminoid|DK –a duplicate of an Original person- reveals emotions convincingly; when following an empirical perspective or when following a theoretical one. The methodology includes the process of acquiring the empirical data and all conclusions are based on the results derived from a questionnaire. The findings suggest that before programming the facial expressions of a Geminoid, the Original should pass through the proposed procedure. Following our recommendations, a believable facial communication is within reach.

Keywords: Social robotics; Geminoid; androids; emotions; facial expressions; emotional health.

1 Introduction

Before the transition from an industrial to a post-industrial society, when the technology-centered view was dominating [1], the words human – computer interaction automatically generated a tag cloud of terms like navigation, manipulation, automation and control. During the post-industrial society, the focus of interaction started to move from the physical machine to the users' world and the interfaces evolved from hardware to social processes [2], [3]. A turn towards human-robot interaction has been observed. Henceforth, during the information era, studies presented a shift of focus towards designing social robots, understanding human behavior and rethinking the role of the machine [4]. Communication theories and cognitive psychology principles were applied to the field of robotics in order to produce as natural interaction as possible and to examine whether the responses from humans were affected by the abilities and the appearance of the robots [5].

The Media Equation communication theory [6] states that human-machine interaction is inherently natural and social, and that the roles of human-human interaction apply also to human-machine interaction. The popularity of the above mentioned theory can be justified easily when realizing that people are used to rely on social and mental models to deconstruct complex behaviors into more familiar, understandable and intuitive forms with which to interact [7]. Despite that fact, following the human perspective to address such issues may not be the most efficient one. Humans are the best known example of emotional interaction, but duplicating human emotional abilities in machines does ensure neither reliability nor perfomativity [8].

In the context of social robotics, when interaction with the user happens in real time, synchronizing the facial gestures and expressions of a robot to the flow of the interaction is essential and eases the communication. For the acceptance of a sociable robot as an equal communication partner, progress is needed at least in the following areas; physical appearance, balanced movements, motions, and gestures, expression and/or perception of emotions, ability to engage in conversation, responsiveness to users and in general ability to process social cues in face to face communication [24], [22]. The more believable and competent it appears, the more users will have the impression of interacting with a human partner rather than just a moving manikin. Even slight improvements to the robots' interface can add credits to the ease of interaction [21].

An issue that is effortlessly rising is "How an android robot can embody emotional facial expressions in respect to the fundamental rules of human affect expression?" [35]. To address this issue, we used the Geminoid|DK android robot [5], [23] which is a teleoperated duplicate of an existing person, the facial expressions of which can be programmed or evaluated by referring to the original person. Reliance only to the facial characteristics of the original person for programming and evaluating the facial expressions of the Geminoid might prove inadequate. In this study, the Geminoid will be tested on the following hypothesis: **The emotional state of the Geminoid|DK when mimicking the Original is equally or more understandable by observers than when following the theoretical approach of Ekman and Friesen in [13].**

We have already presented the train of thought behind this study and in the next sections we will introduce the Geminoid Reality, explain the judgment-base approach, describe the research design methodology for acquiring the empirical data, analyze our findings, validate or not our hypothesis and finally conclude with a discussion about our findings.

2 Geminoids and Related Research

In the last decades, as technology was progressing following the Moore's Law exponential curve [37] and microprocessors were improving in speed and/or performance while decreasing in size, societies were struggling to adapt to these abrupt changes. One after-effect of that fierce technological explosion was the fact that humanity suddenly gained the power to receive and process at the same time, almost parallel to real time, more information than they could physically sense with the natural human sensors. Humanity had the power to build systems, programs, and algorithms that could predict the effects of various changes in the environment, simulate these changes to virtual environments, and be ready to take action in reality. Both the sensors and the results of this enhanced sensing -either of the environment or thyself- could be available on small portable devises. A complex system that encapsulates the same or similar attributes with the one described above, formulates a different kind of reality and is perceived as a different one. Enriching the already known visual reality and objective reality, terms like virtual reality, augmented reality and mixed reality were introduced [11], [18], [26], [28], [30], [36].

2.1 The Geminoid Reality

While virtual reality replaces the real world with a simulated one and augmented reality enhances the real environment with additional virtual information, Geminoid Reality is going one step further. It combines the real environment (robots point of view) with an augmented environment (robots operator point of view) into a new kind of mixed reality with blurry boundaries that is challenging even our perception on visual reality.

In order to be present in the world, it is vital to be ready to engage, to cope and to deal with the world and also to witness events, people and things that are available [27]. Witnessing also requires a record or representation of what has been witnessed. In the Geminoid Reality, the operator of the robot is engaging and coping with the world, dealing with the world and witnessing events, people and things without being physically situated there. His/her presence is substituted by the Geminoid. The operator and the Geminoid form a symbient unity which creates a situation akin to *mirror-touch synesthesia*, a tactile hallucination triggered by observing touch to another person which enables the observer to simulate another's experience by activating the same brain areas that are active when the observer experiences the same emotion or state [12], [15].

The truth about the Geminoid Reality is that it can only be studied in relation to the available technology [20]. Since the field of robotics is still progressing, so the boundaries of Geminoid Reality will keep changing.

2.2 Related Research

Distributed over different kinds of reality, a variety of robotic research has placed emphasis on different aspects of facial expressions, demonstrating the potential in a natural, human like interface. Zoomorphic robots with humanoid facial expression capabilities include MIT’s *Leonardo* [9] and the *iCat* from Philips [10]. In both cases, expression of a range of human emotions is obtained through actuation of the face, often with surprisingly empathic results. Besides the Geminoid series dealt with in this paper, androids with emphasis on facial expressions include the android heads of David Hanson [19] and the *Kansei* head from Meiji University [38]. While the former two robots are often used to obtain and process input from a user as basis for generating an appropriate response, the latter two rely on databases linking emotional expressions to words and phrases.

3 Architecture of Design

3.1 The Judgment-based Approach

The judgment-based approach suggests that when viewing a facial expression, emotion can be recognised entirely out of context, with no other information available. This judgment depends on the judges' past experience of that particular facial expression; either of his/her own face in conjunction with a particular feeling or someone else's face in conjunction with other revealing verbal or non-verbal behavior or in general according to types of activity that have been adaptive in the ancestral past of the species [14], [17].

Both judgment and sign-vehicle based approaches are conceptual types of measurement for studying nonverbal behavior that involve observers, but focus on different phenomena. Judgments about messages are measured in the former and sign-vehicles that convey the message, are measured in the latter. In the message judgment approach the observer is asked to judge whether each subject, who is depicted in a visual input (image or video sequence), is for instance happy or sad, according to the facial expressions that the subject showed. These messages are best represented as dimensions or categories. Observers make inferences about the emotions that underlie the subjects' behavior and for that reason they are referred to as "judges". In the sign-vehicles approach, some or all of the facial expressions and movements would be classified or counted in some fashion, for instance happy subjects lift their cheeks more than the other subjects. Observers describe the surface of behavior by counting how many times the face moves, or how long a movement lasts and are referred to as "coders". As an example, upon seeing a smiling face, an observer with a judgment-based approach would make judgments such as "happy," whereas an observer with a sign-based approach would code the face as having an upward, oblique movement of the lip corners [33].

3.2 Empirical Study

We need three sets of photographs depicting the facial expressions of the Geminoid|DK that correspond to the six basic universally accepted emotions of surprise, fear, disgust, anger, happiness and sadness. In total we need eighteen photographs separated in three equal sets. The first set will be composed of photographs of the Original (O) while expressing the six basic emotions, the second set will be composed of empirically driven (ED) photographs, where the Geminoid will be mimicking the Original and the third set will be composed of theoretically driven (TD) photographs, following the standards of Ekman and Friesen [13].

A dataset of photographs consisting of facial frames of the Original when posing to the camera would only reveal feigned emotions which present significant differences from the prototypical natural ones [25]. In order to depict genuine facial expressions from the Original, all of the six basic emotions had to be triggered. The triggering had to happen naturally, otherwise the results and the whole outcome could be questioned about its reliability. In order to have clear and legible photographs, we decided to subject the Original to a multimodal test on a laptop. The test consisted of a variety of different applications, programs, media, and videos, either online or installed to the Originals' laptop. Each of these was intended to trigger a different emotion. The test was accompanied by a set of instructions explaining the execution order.

The Original was placed in front of his laptop and had one camera pointing at his face, recording the facial expressions, and another one pointing at the screen of the laptop, keeping track of the actions. Afterwards, we had to analyze carefully each of the 44.600 micro-expressions (frames) from the first camera, assign emotions to each of them and finally select six that according to [13] correspond to some universally accepted facial expressions of emotion. The photographs depict one of the many facial expressions an emotion can be revealed. Due to the fact that the triggering was done through a laptop, the eyes of the Original tend to look a bit down (at the screen). Fortunately, this did not affect the outcome of the experiment. Some properties that every photograph shares are that they show the full face under sufficient lighting, are in focus and they are about the same size.

In the figures bellow (Fig.1, Fig.2 and Fig.3), follows a comparison/juxtaposition of the three sets of photographs we took with the facial blueprints of emotions as they are presented in [13]. For the sake of briefness, the figures illustrate only three of the six emotions; Anger, Happiness, and Sadness.

Fig. 1. All facial blueprints for the emotion of Anger (representing Controlled Anger). From left to right: Unmasking the Face (UTF) – TD – ED – O

Fig. 2. All facial blueprints for the emotion of Happiness (representing Full Face Happiness or Intense Happiness). From left to right: UTF – TD – ED – O.

Fig. 3. All facial blueprints for the emotion of Sadness (representing the Lips Down Sad Mouth). From left to right: UTF – TD – ED – O

3.3 Measuring Emotions

As Picard says, even though one has better access to his/her innermost feelings than anyone else, it is still difficult to “recognise” or label the feeling [35]. Through the years, many procedures have been developed, which can be separated into two main categories; the free response measurement and the forced choice response measurement, and the latter consists of two more categories; the dimensional approach and the discrete emotions approach [16], [17].

When emotions are shown as discrete systems, usually the respondents are asked to assess their emotions by a selection from a list of pre-determined emotions or provide feedback on the intensity of an emotion [16]. Following the theories developed by Ekman, we need to be able to select among six different emotions. The research on facial emotions has shown the utility and efficiency of conceiving emotions in discrete states, rather than in dimensions [34]. Dimensional measurements may be most productively applied to emotional experience aggregated across time and to the study of the moods [17]. There appear to be discrete boundaries between the facial expressions of emotion, much as there are perceived boundaries between hues of sound. Facial expressions are perceived categorically and there is accumulating evidence supporting the claim that a discrete system is better applied to momentary experiences of emotion.

3.4 Research Method

We decided to launch an online questionnaire in order to collect feedback. In total we projected 13 videos, each one illustrating a selected facial expression of the Geminoid|DK; 6 with the ED photographs, 6 with the TD photographs and one more video (the first one) that was used as a trial demo.

This questionnaire should not be considered as a test. We don’t intend to train the subjects into recognising emotions, therefore there is no need to force them view each stimulus only once or for a fraction of a second [31], [32]. After all, the facial movements of the Geminoid are mechanical and the change of facial status takes more time than a micro-expression.

This questionnaire should also not be considered as a survey, since there is no need to specify a sampling frame or to ensure sample coverage. According to Ekman, the facial expressions that correspond to the six basic emotions of surprise, fear, disgust, anger, happiness and sadness are considered to be psycho-physiological entities universally accepted [13], [17]. These primary prototypical facial expressions reveal emotions that can be understood universally by people worldwide regardless of gender, nationality, social or economic status and age (except for infants).

The questionnaire is following the judgment-based approach and it can only be described as a judgment and the subjects who respond to the questionnaire can be characterized as judges. For this type of questionnaire, the judges should fulfill two criteria; read and understand the English language, and be at least aware of the android technology in order to avoid "disorientations". Those who follow the Geminoid|DK profile on the Facebook social networking service formulate an acceptable and considerable large group of people that are aware of the existence of android robots, are familiar at least with the Geminoid technology, have access to an computer with internet connection, and know how to communicate in English.

An online questionnaire can provide more honest answers than if choosing another method, as respondents feel that their privacy is not violated [29]. Due to the medium that the questionnaire was launched, and due to time and cost constraints, the judges were random people who had access to the link.

Instead of using static photographs as stimuli, we decided to use videos. Most facial expressions of emotions during a conversation last between 500ms and 2.5s [32]. Displaying a short video with the desired facial expression was considered to be a more reliable option. The video would last three seconds. During the 1st and 3rd second the video would be blank, and during the 2nd second it would display the frame of the facial expression. The whole procedure reminds the blinking of the eye. We considered thoughtful to add another frame at the beginning of the video lasting 4 seconds warning viewers about the briefness of the video. Judges could view the videos as many times as they wanted and they were prompted to answer the question "*What emotion do you think the face in the video is showing?*" by selecting from a list with the pre-determined six basic emotions on a two-point intensity scale; either the emotion existed in the facial expression or not.

4 Analysis of the results

The online questionnaire was reached by many people worldwide (678 unique visitors), but only 50 people (34 females and 16 males) actually filled and submitted it. The overrepresentation of female respondents was anticipated [29]. The judges were located mainly in Europe, except for two coming from America, and belonged mainly to the age group of 21-30 years old (one was less than 15, nine belonged to the group of 16-21, thirty belonged to the group of 21-30, eight belonged to the group of 31-40, one to the group of 41-50 and one belonged to the group 51-60).

The results for the Surprise videos indicated that the strong majority of the judges named the emotion of Surprise as the dominant one, understanding our intention. The judgments for the Surprise emotion in the emotionally driven videos (EDV) outnumber the ones of the theoretically driven videos (TDV) (45 and 39 respectively), validating our hypothesis. The second most dominant emotion for the EDV was Happiness with 24 votes and for the TDV was Anger with 24 votes.

In the Fear videos, the judgments favored another emotion than the one we intended to show. Instead of Fear, both videos revealed mainly the emotion of Surprise (49 judgments for the EDV and 44 for the TDV). This outcome suggests that the Geminoid is a substitute expressor for Fear. The emotion of Surprise is quite pronounced, coloring the whole facial expression. However, the emotion of Fear was better understood in the EMV than in the TDV (32 against 26 judgments), so the hypothesis is satisfied.

For the Disgust videos, there was no agreement among the judges about an emotion. No more than a third of the judges gave any one judgment**,** excluding only the emotion of sadness (28 judgments) for the EDV and happiness for the TDV (17 judgments). This result suggests that the Geminoid|DK might be a withholder. The Disgust judgments were equally distributed (5 each), validating again the hypothesis.

Both of the Anger videos were judged as we indented to (34 judgments for the EDV and 14 for the TDV), which confirm our hypothesis. The emotion that came second in the judgments was Disgust with 11votes (EDV) and Happiness with also 11 votes (TDV).

The Happiness videos were also understood (45 judgments for the EDV and 44 for the TDV), but the judges distinguished also the emotion of Surprise (28 judgments for the EDV and 30 for the TDV). This outcome suggests that the Geminoid|DK is a substitute expressor to a slight degree that is not so evident. One more time the hypothesis was confirmed.

Results for the Sadness videos revealed that there was almost an even split between the intended emotion of Sadness and another emotion – that of Anger. This outcome suggests that the Geminoid|DK might be an Anger-for-Sadness substitute expressor, as Anger was the second most dominant emotion at the EDV with 9 judgments and the most dominant at the TDV with 19. Observers' judgments matched with what we were expecting only in the EDV (24 judgments in the EDV against 12 in the TDV); therefore the hypothesis is confirmed.

5 Conclusion

We have presented the current status of the Geminoid technology and what it needs to become Geminoid Reality. Apart from an adaptive interface able to communicate with and understand the surrounding environment, it also needs to have believable characteristics and be able to actively engage in interaction. The process of finding the ways a Geminoid can embody emotional facial expressions in respect to the fundamental rules of human affect expression was based on the work of Ekman, only shifted to the robotic field.

The results of the questionnaire revealed an incapability of the Geminoid to reproduce the emotions of Fear and Disgust. Our proposal concerns the next version of the Geminoid series. We believe that installing actuators to the facial areas of the *levator labii superioris* /*alaeque nasi* (nose wrinkle), *levator labii superioris/caput* *infraorbitalis* (upper lip raiser), *depressor anguli oris (triangularis)* (lip corner depressor), *incisivii labii superioris* and *incisivii labii inferioris* (lip puckerer) and *orbicularis oris* (lip tightener) will make expressions more human-like. Another addition that would reveal more natural facial expressions could be to make the already installed actuators operate independently (i.e., to lift just one eye-brow).

Lastly, we have proven that the emotional state of the Geminoid|DK is equally or more understandable when mimicking its' Original than when following the theoretical approach of [13]. This finding suggests that before programming the facial expressions of a Geminoid, the Original should pass through a similar procedure. Following our recommendations, a believable facial communication is within reach.

References

1. Krippendorff, K.: The semantic turn - A new foundation for design. CRC Press, Taylor & Francis Group (2006)
2. Grudin, J.: The computer reaches out: The historical continuity of interface design. In: Chew, J.C., Whiteside., J. (eds.) Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people (CHI '90), ACM, NY (1990)
3. Dourish, P.: Where The Action Is: The Foundations of Embodied Interaction. MIT Press (2001)
4. Breazeal, C.: Social Interactions in HRI: The Robot View. In: IEEE Transactions on Systems Man ,Cybernetics and Systems-Part C, vol. 34, issue 2, pp. 181-186, IEEE Press, new York (2004)
5. Nishio, S., Ishiguro, H., Hagita, N.: Geminoid: Teleoperated Android of an Existing Person. In: Filho, A.C.D.P. (ed.) Humanoid Robots:New Developments, pp. 343-352, Vienna (2007)
6. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. University Chicago Press (1996)
7. Breazeal, C.: Toward sociable robots. In: Robotics and Autonomous Systems, vol. 42, issues 3–4, pp. 167-175. Elsevier (2003)
8. Picard, R.W., Klein, J.: Computers that Recognise and Respond to User Emotion: Theoretical and Practical Implications. In: Interacting with Computers, vol. 14, issue 2, pp. 141-169. Elsevier (2002)
9. Personal Robotic Group-Leonardo, http://robotic.media.mit.edu/
10. Philips-iCat, http://www.research.philips.com/technologies/projects/robotics/
11. Hoffman, D.D.: Chapter 2: The Construction of Visual Reality. In: Blom, J.D., Sommer, I.E.C. (eds.) Hallucinations: Research and Practice, pp.7-15. Springer US (2012)
12. Blom, J.D., Sommer, I.E.C.: Chapter 13, Hallucinations of Bodily Sensation. In: Blom, J.D., Sommer, I.E.C. (eds.) Hallucinations of Bodily Sensation , Part 2, pp. 157-169. Springer US (2012)
13. Ekman, P., Friesen, W.V.: Unmasking the Face – A guide to recognising emotions from facial expressions. Malor Books, Cambridge (2003)
14. Cohn, J.F.: Foundations of Human Computing: Facial Expression and Emotion. In: Proceedings of the 8th international conference on Multimodal interfaces (ICMI'06), pp. 233-238. ACM, New York (2006)
15. Banissy, M.J., Ward, J.: Mirror-touch synesthesia is linked with empathy. In: Nature Neuroscience , vol. 10, pp. 815 – 816. NPG (2007)
16. Scherer, K.R.: What are emotions? And how can they be measured? In: Social Science Information, vol. 44, issue 4, pp. 695-729. SAGE Publications (2005)

17. Keltner, D., Ekman, P.: Facial expression of emotion. In: Lewis, M., Jones, J.H. (eds.) Handbook of emotions, 2nd edition, pp. 236-248, Guilford Publications, New York (2000)
18. Behrendt, R.P.: Consciousness, Memory, and Hallucinations. In: Blom, J.D., Sommer, I.E.C. (eds.) Hallucinations: Research and Practice, part 1, pp. 17-31. Springer US (2012)
19. Hanson Robotics, http://hansonrobotics.wordpress.com/
20. Thrun, S.: Towards A Framework for Human-Robot Interaction. In: Journal Human-Computer Interaction, vol. 19, issue 1, pp. 9-24. Taylor & Francis (2004)
21. Takano, E., Matsumoto, Y., Nakamura, Y., Ishiguro, H., Sugamoto, K.: The Psychological Effects of Attendance of an Android on Communication. In: Experimental Robotics, Springer Tracts in Advanced Robotics, vol. 54, pp. 221-228. Springer, Berlin / Heidelberg (2009)
22. Becker-Asano, C., Wachsmuth, I.: Affective computing with primary and secondary emotions in a virtual human. In: Auton Agent Multi-Agent Syst , vol. 20, issue 1, pp. 32–49. Springer, Netherlands (2010)
23. Geminoid|DK, http://www.geminoid.dk
24. Shimada, M., Yoshikawa, Y., Asada, M., Saiwaki, N., Ishiguro, H.: Effects of Observing Eye Contact between a Robot and Another Person. In: International Journal of Social Robotics, vol. 3, issue 2, pp. 143–154. Springer (2011)
25. Padgett, C., Cottrell, G.: Representing face images for emotion classification. In: Mozer, M., Jordan M., Petsche, T (eds.) Advances in Neural Information Processing Systems, vol. 9, Cambridge, MA. MIT Press (1997)
26. Taylor, J.: The Emerging Geographies of Virtual Worlds. In: Geographical Review, vol. 87, issue 2, Cyberspace and Geographical Space, pp. 172-192. American Geographical Society (1997)
27. Turner, P.: An everyday account of witnessing. In: AI & Society: Special issue: Witnessed presence, vol.27, issue 1, pp. 5–12. Springer-Verlag, London (2012)
28. Holz, T., Campbell, A.G., O'Hare, G.M.P., Stafford, J.W., Martin. A., Dragone, M.: MiRA—Mixed Reality Agents. In: International. Journal of Human – Computer Studies, vol. 69, issue 4, pp. 251–268. Academic Press, (2011)
29. The American Statistical Association, http://www.amstat.org/
30. Cheok, A.D., Haller, M., Fernando, O. N. N., Wijesena, J. P.: Mixed Reality Entertainment and Art. In: The International Journal of Virtual Reality, vol. 8, issue 2, pp. 83-90 (2009)
31. Boucher, J.D., Ekman, P.: Facial Areas and Emotional Information. In: Journal of Communication, vol. 25, issue 2, pp. 21-29 (1975)
32. Ekman, P.: Emotions Revealed-Understanding Faces and Feelings. Phoenix, London (2003)
33. Cohn, J.F., Ekman, P.: Measuring facial action by manual coding, facial EMG, and automatic facial image analysis. In: J. A. Harrigan, J.A., Rosenthal, R., Scherer, K. (eds.) Handbook of nonverbal behavior research methods in the affective sciences, pp. 9 - 64 (2005)
34. Ekman, P.: Facial expression and emotion. In: American Psychologist, vol. 48, issue 4, pp. 384-392 (1993)
35. Picard, R. W.: Affective Computing: Challenges, In: International Journal of Human-Computer Studies, Vol. 59, Issues 1–2, pp. 55–64. Academic Press (2003)
36. Carmigniani J., Furht, B.: Augmented Reality: An Overview. In: Furht, B. (ed.) Handbook of Augmented Reality, Part 1, pp. 3-46. Springer, New York (2011)
37. Vardi M.Y.: Is Moore's Party Over? In: Communications of the ACM , vol. 54, No. 11, pp. 5. ACM, New York (2011)
38. Takeno, J., Mori, K., Naito, Y.: Robot Consciousness and Representation of Facial Expressions. In: 3rd International Conference on Sensing Technology (ICST 2008), pp. 569-574. IEEE Press (2008)