Response Times and Accuracy when judging emotions in realistic and virtual faces

Human Perception for Information Technology

Oliver Shin oliversh@kth.se, Priyanka Syal syal@kth.se, Sruti Bhattacharya srutib@kth.se,
Wenjia Chen wenjiac@kth.se

Abstract

This short study investigates the perception of emotional expressions in realistic vs virtual faces with regard to response time and identification accuracy. Stimuli were chosen from two separate datasets, one containing realistic faces and one containing virtual looking ones.

The OpenSesame software was used to design an interactive program that measured the response time and logged the identifications made by each of the 12 participants. The data gathered from these trials were condensed into a series of mean average values and compared. A general indication based on the results suggests that happiness in realistic faces was the quickest and most accurately identified stimuli type during the experiment.

Some challenges were encountered during the development process of the experiment interface resulting in some aesthetic and functional drawbacks. These were mitigated as much as possible through use of practice blocks allowing the participants to familiarize themselves with the interface before starting the main experiment.

Conclusively, it is worth noting that no statistical significance was attained due to the relatively few study participants. Viewing the results as qualitative however still offers some indication that supports previous research claims such as happiness being a relatively easy emotion to judge, both for realistic and virtual faces.

1. Introduction

Facial expression is commonly regarded as a fundamental and effective way to display basic emotional states, which is incredibly important for human social interaction. Though emotional expression is affected by the culture environment of the expresser and observer, some of them can be signified universally [1].

For entertainment and academic purposes, vivid virtual characters with emotional expressions were also created [2]. However,

the recognition of expression is not always accurate for most observers, whether they are identifying real faces [3-5] or virtual ones [4,5]. There are several factors affecting. Palermo and Coltheart's study shows the emotion type and face sex in photographs influences recognition rates [3]. Previous experiments [4,5] revealed that age impacted ability to recognize facial expressions.

Since people are moving more and more toward communication through digital channels, online interactions have seen an incredible surge in widespread use. With the latest technology, users can show not only real faces but also animated avatars with facial expression in video calls and webcasts. Accessing emotion from natural and virtual faces may become the same important in the future. This situation inspires us to advance knowledge in perception of emotional facial expressions.

The present study aims to investigate recognition ability in humans when distinguishing emotional expressions on natural vs virtual faces. Accuracy rate and response time can be considered as two elements that are relevant and suitable for measure. Following questions were suggested for the research:

Is there a difference in response time when identifying emotional expressions in realistic vs virtual images?

Will the accuracy of the identification of virtual face expressions be as reliable as that of a realistic photographed face?

To solve these problems, a perceptual experiment is built with the help of OpenSesame [10] software to show testing slides and record participants' responses.

2. Background

2.1 Universal Facial Expressions

According to Ekman [1], particular emotions are associated with distinctive facial muscular movements among multiple cultures, such as: anger, disgust, fear, happiness, sadness, and surprise. Though the theory of basic emotions and facial expressions keeps developing in continuous researches [6], many studies with relation to emotional expressions are based on the basic emotions mentioned above that can be encoded via universal facial expressions, including a neutral emotionless expression [3-5,7]. We decided to choose a subset of universal facial expressions for conducting our experiment.

2.2 Facial Expression Datasets

Realistic and virtual faces for this experiment were selected from two databases separately: the Karolinska Directed Emotional Faces (KDEF) [8], and

UIBVFED virtual facial expression dataset [9]. Both datasets provide representations of neutral and all the six basic expressions.

2.2.1 KDEF Dataset

KDEF datasets consist of 4900 photographs from 70 amateur actors (35 males, 35 females) displaying 7 different facial expressions of basic emotions. Their age ranges from 20 to 30 years. Each expression is pictured twice from 5 different angles. Since this a standardized dataset used across different research studies and was easily available, we used a subset of images taken from the straight angle for the purpose of our study.

Figure 1 - Sample stimuli from KDEF

2.2.2. UIBVFED Dataset

UIBVFED dataset consists of 640 pictures from 20 virtual characters (10 males, 10 females) each presenting 32 facial expressions. The avatars were created from various ethnicities with ages between 20 and 80 years old. Images are categorized under the 6 universal expressions (and 1 neutral).

Figure 2 - Sample stimuli from UIBVFED

We chose a subset of these images corresponding to neutral, happiness (open mouthed smiling expression), sadness (nearly crying expression) and surprise for our main experiment.

2.3 OpenSesame

OpenSesame is a ‘free, opensource, and cross-platform’ software designed for users to build experiments, especially those relating to the field of psychology, neuroscience, and experimental economy [10]. The application has a streamlined graphic interface, making it friendly for

beginners. It also supports complex experiments via its compatibility of Python coding and plug-in functionality.

OpenSesame is frequently used in experiments for conducting academic experiments. Since it performs well in handling the task of recording responses and response times and supports various modes of ordering and randomization we chose it to create our experiment. The latest stable version (3.3.3) on MacOS was used to create the main experiment.

3. Method

Our perception experiment was created by combining various existing tools and studies using OpenSesame, 3D models, photoshop. We also used existing datasets for Real and virtual faces for this experiment. We used the **Karolinska Directed Emotional Faces (KDEF)** [8] for real faces , and **UIBVFED** virtual facial expression dataset [9] for virtual faces. We surveyed a total of 21 users in our prestudy over a week and conducted the experiment with the total of 12 users over the course of five days.

We used a subset of images taken from the straight angle in the KDEF and a subset of images from UIBVFED dataset corresponding to neutral, happiness (open mouthed smiling expression), sadness (nearly crying expression) and surprise for our main experiment (Table 1 in Appendix).

3.1 Prestudy

A prestudy was conducted to gather responses from test subjects using accidental sampling of unknown gender, age and occupation. Participants performed the prestudy by answering a google form survey showing a virtual expression along with the realistic photographic counterpart of the same expression. For each comparison, participants were asked to rate how similar the two images are w.r.t to their expressions using the three-step similarity scale: 1 - not similar, 2 -similar and 3 - very similar. The ambition was to identify potential outlier stimuli pairs that participants do not identify as showing the same emotion. A total of 21 participants took the initial round of prestudy.

3.2 Selection and design of stimuli

24 stimuli pairs showed results where over half of the participants categorized them as “similar”. The remaining 4 pairs of stimuli were identified as being too dissimilar. Disgust - Female, Disgust - Male, Anger - Female and Sadness - Male all had a response percentage majority(more than 50%) for the option “not similar”. This prompted a second round of the same survey with new stimuli pairs in an attempt to identify pairs of higher similarity. This second round was answered by a total of 11 participants. Ideally this process would repeat until a satisfiable level of similarity

was achieved for all stimuli pairs. However, due to time constraints this could not be completed. Furthermore, the realistic photograph stimuli gathered from the **KDEF** all had a reddish tint. Color association can potentially contribute to evoking certain emotions and might also differ depending on cultural context[14]. With this in mind, all **KDEF** images were converted to gray scale using photoshop to limit the emotional influence of color. Additionally, we realized that we would need multiple stimuli for each category of expression and stimuli type that we wanted to assess. This was to validate and increase the accuracy of our results. Since we had decided that we will include both genders for both real and virtual faces we had four categories of stimuli type - virtual female, virtual male, real female and real male for each expression. To get higher accuracy of the results and minimize the variability and bias, we decided to have three different stimuli for each type for each expression. We wanted to limit the time of the experiment to make sure that we are getting the maximum of the user's attention span. Additionally, based on the previous studies [16] and our prestudy results where people are we realized disgust, fear and anger were very difficult to depict in 3D avatar. Furthermore, based on the limitation of our dataset and within the constraint of limiting our experiment we decided to focus on only four emotions where we had sufficient stimuli for depicting expressions. Based on this, we chose to select three stimuli of each stimuli type for the expression of neutral, sadness, happiness and surprise.

Figure 3 - A sample set of stimuli chosen for the experiment

3.3 Experiment design and Procedure

The OpenSesame experiment we developed starts by asking consent from the user. We used google forms to gather the age and gender of the users. Once consent had been received, the user was prompted to perform the practice block. We also explained the instructions to the users. The practice block consisted of 4 practice slides. Each slide consisted of a picture of the stimuli with the question “What expression is the character making?” with the seven emotions as the options.

The purpose of the practice block was to help the user get familiar with the experiment setup. Once the practice block was over we showed the user a slide saying that the main experiment is starting. We divided the main experiment into three blocks. Each block had four kinds of stimuli

- one male 3D avatar, one female 3D avatar, one male photograph, one female photograph exhibiting four emotional expressions (happy, sad, neutral and surprise). We used Opensesame's randomization algorithm to randomize the order of images in each block which is internally implemented using python's pseudo-random algorithm. To prevent two stimuli of the same type appearing with the same expression, we made sure that each block had only one instance of the same type of stimuli per expression. Hence we had a total of 16 slides in one block and 48 slides total which we evaluated per user. We also added a slide between each block where the user could take a break if they wanted. The total experiment had total 58 slides including the practice, instructions, consent, and slide indicating end of each block and the end of experiment.

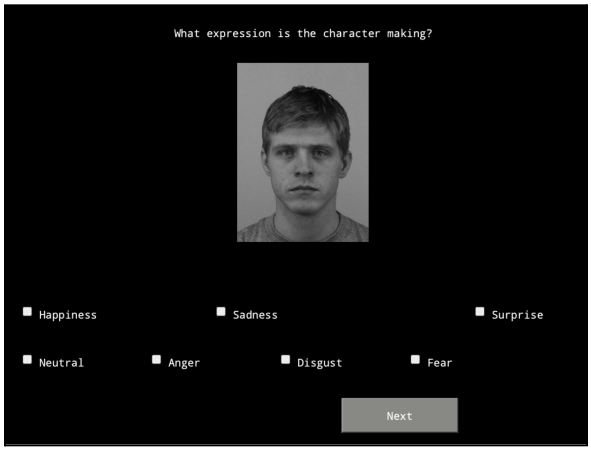

Figure 4 - Sample of the slide shown to the user in the experiment

The user feedback was saved using OpenSesame that saves the experiment results into a CSV (Comma Separated Values) file. The file was anonymized by only saving the participant number and contained the random ordering of the stimuli for every participant. It also contained the response time and perceived emotion by the participant for every stimuli including the practice session.

3.3.1 Pilot Study

We conducted an initial pilot study with our team members. Using the feedback from the initial run, we thought of adding a practice block of 4 slides where the users would get comfortable with the interface.

3.3.2 Main Study and Experiment Execution

We conducted the experiments at Multimedia Lab, KTH, and some of them at Fylke. The experiments were conducted taking into account safety measures for COVID. We provided the participants with sanitizing wipes before and after the experiment. We ensured that the equipment was also sanitized before the start and the end of each experiment. We followed social distancing norms while experimenting and allowed only one participant at a time in the lab room.

Initially, 18 people booked the 15-minute slots to participate in the experiment through

our doodle link. But, finally, only 12 people took the experiment due to various reasons. Each test participant sat on a chair in front of the system (We used a Mac Desktop at Multistudio lab and a MacBook Pro laptop at Fylke) with an average eye-display distance of 80 centimeters. Participants were given a brief verbal introduction to the experiment by a project member and initial guidance about what to do in the experiment.

3.3.3 Data Analysis

The CSV files were aggregated and the experiment results were analyzed using Excel. The method of determining the average values used in the tables was attained using as an arithmetic mean value.

3.3.4 Measures

Response time (RT) is the time measured in milliseconds until the first mouse click interaction with the experiment interface.

Accuracy, referring to the percentage of correctly recognized expressions of each emotion in the experiment.

4. Results

The following results are divided into two parts. The first presenting the response time required to identify an emotion based on a face stimuli. The second one presents the accuracy of emotional identification.

4.1 Response time

The total average response time was calculated by taking the mean across all the average times per subject. Furthermore additional results can be retrieved by comparing the average response time per emotion as shown below.

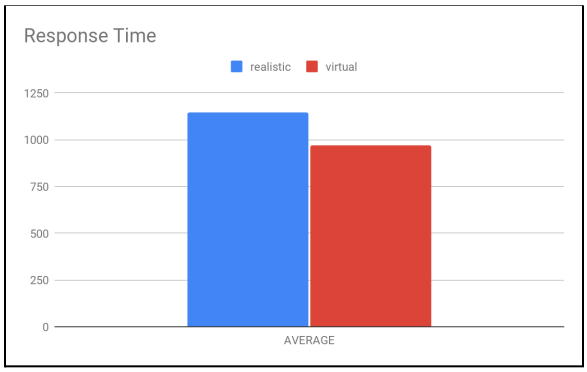

Figure 5 - total average response time for virtual and realistic faces

As illustrated by Figure 1 the average response time for identifying emotions were fairly similar between realistic and virtual faces. The realistic faces having a slightly slower average time before identification.

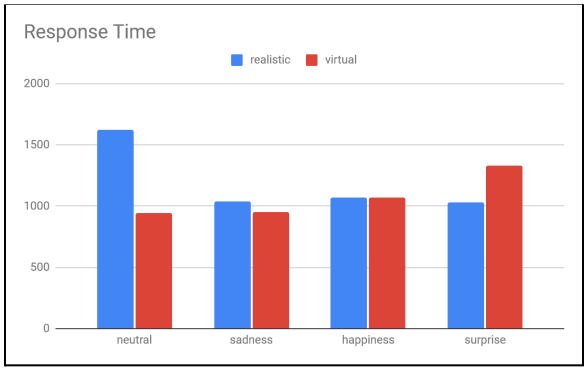

Figure 6 - Average response time per emotion for virtual and realistic faces of correctly identified emotions

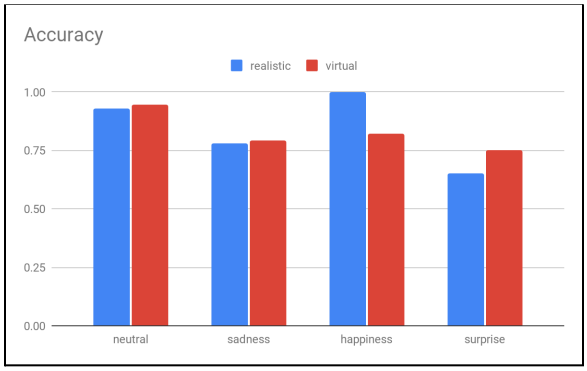

Figure 7 - Average accuracy of identification percentage of virtual and realistic faces

The average identification times are similar for the emotions sadness and happiness while neutral and surprise exhibit some deviances. A neutral emotional expression in realistic looking stimuli seems to have taken the longest to identify of all. To a lesser extent virtual faces showing surprise also seemed to take slightly longer to identify compared to happiness or sadness.

4.2 Accuracy

The following figure shows the accuracy of identification.

Certain emotions appear to have been harder to identify than others. The stimuli used in the experiment have performed differently with some generating less accurate results than others. Realistic photos of a happy face stimuli shows a 100% accuracy rate meaning all subjects correctly identified these stimuli correctly. The most difficult to identify seems to have been surprise. This holds true for both virtual and realistic faces.

5. Discussion and Analysis

It is worth noting that human emotional perception is largely affected by the context in which the stimuli is experienced[11]. The results gathered in this study thus seem to indicate that the emotional expression of happiness(also perhaps others to a lesser extent) might be more universally identifiable at least in this particular context. “Context” however, can be vast and

complex to fathom. Subtle changes to the stimuli as the angle of observance might have a determinant effect on how quickly we can associate the stimuli with a recognizable emotion[13]. Other important aspects of context as mentioned previously in the experiment design section include color variance which can convey certain emotional connotations interfering with the intended emotion[14]. Furthermore, body language and physical context of surroundings may have a causal effect on the way we experience a situation and especially when dealing with facial stimuli that do not necessarily exhibit large physical dissimilarities. An example described by Munn is when compartmentalizing the image of an athlete to seeing only the facial expression. Munn found in their study that this was judged as sorrow, but when seeing the entire body in its physical context the stimuli was identified as determination[15].

Finally, when looking over the results of the study there does not appear to be any conclusive or significant differences for identifying happy, sad, neutral and surprised expressions between virtual and realistic looking faces. The differences that do appear might as well be attributed to the design of the stimuli as indicated by the prestudy. Some of the stimuli were considered by the participants to not accurately reflect their intended emotion.

5.1 Delimitations

Primary delimitations worth mentioning is the participant scope and stimuli design. Due to the limitation of time and resources, the study was based on the result of only 12 participants which does not give enough quantitative raw data in order to achieve statistical significance. However, the qualitative opinions are still worth considering as emotional perception partly subjective. Cultural and personal context play a role in how we interpret emotional stimuli[12]. The stimuli design was unfortunately not fully realized according to the initial plan due to time constraints. Ideally there would have been further iteration on investigating the stimuli similarity to their intended emotion. Additionally, we used the opensesame's inbuilt randomization for ordering of the stimuli. If we had more time, we could have explored other methods of ordering and ensured that the stimuli is more randomized.

Other points worth considering going forward is the cultural diversity of the faces. Especially for the Karolinska Directed Emotional Faces (KDEF) [8] which consisted almost completely of persons with a white caucasian complexion.

Additionally, we only used the front faces from our dataset and did not consider perception of emotions if the stimuli was viewed from different angles. Furthermore,

for this experiment we have assumed that our participants were looking straight and have not taken users eye gaze and head position into account.

5.2 Project assessment

The experiment software that we developed turned out fairly good but with a few quirks and aesthetical drawbacks. We underestimated the amount of work that would be required to use OpenSesame which led to unexpected workload. Finding persons to participate in our study was fairly challenging and with regard to the ongoing covid-19 pandemic, adherence to quarantine guidelines made it difficult to organize the experiment in practice. Team collaboration was also fairly challenging as most meetings and planning sessions had to be carried out over voice call. This often resulted in miscommunication and lack of transparency.

A main takeaway from this project is awareness of the challenge presented when designing a perception stimuli. There are an abundance of factors that can interfere with the stimuli's ability to evoke a certain emotion, ranging from ordering bias to stimuli color hue.

6. Conclusion

The premise for this study was to explore whether there is a difference in response time when identifying emotional expressions in realistic vs virtual images? Also, Will the accuracy of the identification of virtual face expressions be as reliable and accurate as that of a realistic photograph face? Conclusively we can summarize our findings as a qualitative indication that there is not a great difference in average response time when comparing the identification of virtual vs realistic face stimuli, except for neutral face expressions. The difference between realistic and neutral identification was approximately 700 milliseconds(Figure 6). The accuracy of identification seems to line up with results from previous studies suggesting that happiness is the easiest emotion to identify when observing a realistic face from a frontal angle[3,4]. However, the results of this study point towards happiness of a virtual face being about as hard to identify correctly as neutral, surprise and sadness. One possible explanation for this can be that identification of emotions from face stimuli can have different levels of contextual dependency[15].

7. Further Development

Apart from the stimuli, eye gaze and head position can also play a role in perception of emotion and it would be interesting to explore that area. Additionally, the study could also be extended to measure perception of the emotion if the stimuli was shown in different angles. Furthermore, it would be interesting to evaluate, create and

use stimuli displaying different intensities of emotions or according to other models of emotions and expressions like the Geneva Emotional Wheel [17].

8. References

1. Ekman P. Universals and cultural differences in facial expressions of emotion. Nebraska Symp Motiv. Published online 1971.

2. Vinayagamoorthy V, Gillies M, Steed A, et al. Building Expression into Virtual Characters: State of The Art Report. Published online 2006.

3. Palermo R, Coltheart M. Photographs of facial expression: Accuracy, response times, and ratings of intensity. Behav Res Methods, Instruments, Comput. 2004;36(4):634-638. doi:10.3758/BF03206544

4. Dyck M, Winbeck M, Leiberg S, Chen Y, Gur RC, Mathiak K. Recognition profile of emotions in natural and virtual faces. PLoS One. Published online 2008. doi:10.1371/journal.pone.0003628

5. Beer JM, Smarr C-A, Fisk AD, Rogers WA. Younger and older users′ recognition of virtual agent facial expressions. Int J Hum Comput Stud. 2015;75:1-20. doi:10.1016/j.ijhcs.2014.11.005

6. Ekman P. Basic Emotions. In: Handbook of Cognition and Emotion. ; 2005:45-60. doi:10.1002/0470013494.ch3

7. McDonnell R, Jörg S, McHugh J, Newell F, O'Sullivan C. Evaluating the emotional content of human motions on real and virtual characters. APGV 2008 - Proc Symp Appl Percept Graph Vis. Published online 2008:67-73. doi:10.1145/1394281.1394294

8. Lundqvist D, Flykt A, Ohman A. The Karolinska directed emotional faces (KDEF). CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet. Published 1998. https://www.emotionlab.se/kdef/

9. Oliver MM, Alcover EA. UIBVFED: Virtual facial expression dataset. PLoS One. 2020;15(4). doi:10.1371/journal.pone.0231266

10. Mathôt S, Schreij D, Theeuwes J. OpenSesame: An open-source, graphical experiment builder for the social sciences. Behav Res Methods. Published online 2012. doi:10.3758/s13428-011-0168-7

11. Hess, Ursula, Blaison, Christophe, & Kafetsios, Konstantinos. (2016). Judging Facial Emotion Expressions in Context: The Influence of Culture and Self-Construal Orientation. Journal of Nonverbal Behavior, 40(1), 55-64.

12. Masuda, Takahiko, Ellsworth, Phoebe C, Mesquita, Batja, Leu, Janxin, Tanida,

Shigehito, & Veerdonk, Ellen Van de. (2008). Placing the Face in Context. Journal of Personality and Social Psychology, 94(3), 365-381.

13. MICHIKO, H. (1992). Perception of facial expressions: Effect of facial angle on emotion category effect. Tohoku Psychologica Folia, 52, 33-36.

14. Wilms, Lisa, Wilms, Lisa, Oberfeld, Daniel, & Oberfeld, Daniel. (2018). Color and emotion: Effects of hue, saturation, and brightness. Psychological Research, 82(5), 896-914.

15. Munn, N. L. (1940). The effect of knowledge of the situation upon judgment of emotion from facial expressions. The Journal of Abnormal and Social Psychology, 35(3), 324–338. https://doi.org/10.1037/h0063680

16. Noël, S., Dumoulin, S., & Lindgaard, G. (2009, August). Interpreting human and avatar facial expressions. In IFIP Conference on Human-Computer Interaction (pp. 98-110). Springer, Berlin, Heidelberg.

17. Klaus R Scherer. 2005. What are emotions? And how can they be measured?Social science information 44, 4 (2005), 695–729.

Appendix

Table 1

Stimuli Type	Happiness	Sadness	Neutral	Surprise
KDEF - male	AM07HAS.jpg	AM01SAS.jpg	AM10NES.jpg	AM03SUS.jpg
	AM08HAS.jpg	AM28SAS.jpg	AM23NES.jpg	AM06SUS.jpg
	AM12HAS.jpg	AM31SAS.jpg	AM29NES.jpg	BM05SUS.jpg
KDEF - female	AF07HAS.jpg	AF02SAS.jpg	AF01NES.jpg	AF07SUS.jpg

	AF18HAS.jpg BF24HAS.jpg	AF03SAS.jpg AF15SAS.jpg	AF04NES.jpg AF06NES.jpg	AF20SUS.jpg AF26SUS.jpg
UIBVFED - male	Mateu_SmilingOpenMouthed.png Pere_SmilingOpenMouthed.png Tomeu_SmilingOpenMouthed.png	Allan_Miserable.png Dave_NearlyCrying.png Kurt_NearlyCrying.png	Allan_Neutral.png Jose_Neutral.png Mateu_Neutral.png	Dave_Surprise.png Lluis_Surprise.png Ramon_Surprise.png
UIBFED - female	Maria_SmilingOpenMouthed.png Alicia_SmilingOpenMouthed.png Sebastiana_SmilingOpenMouthed.png	Cristina_NearlyCrying.png Isabel_NearlyCrying.png Wanda_NearlyCrying.png	Alicia_Neutral.png Cristina_Neutral.png Sebastiana_Neutral.png	Catalina_Surprise.png Francisca_Surprise.png Wanda_Surprise.png

Prestudy results

https://docs.google.com/spreadsheets/d/1qzozdZMBnCy94fi07Trfivol6vyPNSRrU4mjlBhLSwg/edit?usp=sharing