Discriminating Real and Posed Smiles: Human and Avatar Smiles

Md Zakir Hossain **Tom Gedeon**
The Australian National University RSCS, ANU, Canberra 2601, Australia
{zakir.hossain, tom}@anu.edu.au

ABSTRACT

This study addresses the question whether untrained and unaided observers can discriminate between real and posed smiles from different sets of smiling videos. Observers were shown smiling face videos, either in single or paired, to rate each of them on a number of scales intended to assess perceived real and posed smiles. We implemented four experiments where single smiles were shown in most cases, and paired smiles were shown in the third experiment. We found that observers are more accurate in response to paired videos (72.9%) compared to single videos (61.7% and 60.2%). Performance is increased when voting is introduced, but we need 11 to 13 observers in single smiles and 13 to 15 in paired smile discrimination. We found that female observers are more accurate compared to male observers. On testing with 8 'smiling' virtual avatars we found only one was rated a real smile. This work will have significant impact on judgement of genuineness of smiles from avatars / automated virtual assistants, in that untrained individuals using normal human abilities can estimate how real the virtual smiles will seem for the human users to whom they are directed.

CCS CONCEPTS

• **Human-centered computing** → **Interaction design** → Scenario-based design; • **Information systems** → **Information retrieval** → Users and interactive retrieval

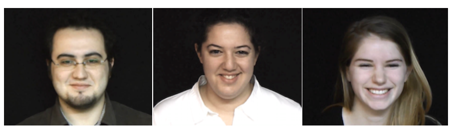

Figure 1: Sample frames: real smiles

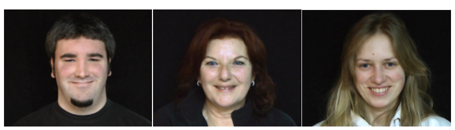

Figure 2: Sample frames: posed smiles

Figure 3: Sample frames: avatar smiles

OzCHI '17, November 28-December 1, 2017, Brisbane, QLD, Australia

ACM ISBN 978-1-4503-5379-3/17/11
https://doi.org/10.1145/3152771.3156179

Table 1: Experiment demographics; M=Male, F=Female, Observers = Obs.

	Obs.			Age	
	M	F	Total	Mean	±Std.
E1	6	5	11	31.6	5.0
E2	18	13	31	28.8	6.8
E3	17	14	31	21.9	3.0
E4	7	8	15	33.6	4.5

Table 2: Obs. VR (%) in E1

O1	*O2*	*O3*	*O4*	*O5*	*O6*
54.2	50.0	70.8	62.5	66.7	66.7
O7	*O8*	*O9*	*O10*	*O11*	***Avg.***
70.8	58.3	66.7	58.3	54.2	**61.7**

Table 3: Obs. VR (%) in E2

O1	*O2*	*O3*	*O4*	*O5*	*O6*
65	40	75	75	45	65
O7	*O8*	*O9*	*O10*	*O11*	*O12*
60	75	50	55	50	55
O13	*O14*	*O15*	*O16*	*O17*	*O18*
60	75	65	55	70	45
O19	*O20*	*O21*	*O22*	*O23*	*O24*
45	60	55	45	70	60
O25	*O26*	*O27*	*O28*	*O29*	*O30*
85	85	50	85	40	60
O31	***Avg.***				
60	**60.2**				

KEYWORDS

Observers, Smile Videos, Verbal Response, Real Smile, Posed Smile, Virtual Reality, Avatar

ACM Reference format:

Md Zakir Hossain, and Tom Gedeon, 2017, Discriminating Real and Posed Smiles: Human and Avatar Smiles. In *Proceedings of the 29th Australian Conference on Human-Computer Interaction* (OzCHI '17), Brisbane, QLD, Australia, November 2017, 6 pages. https://doi.org/10.1145/3152771.3156179

1 INTRODUCTION

Facial expressions are intrinsically linked to the human emotional information system that carries critical social signals and incites accurate, clear, and immediate recognition [4]. This information is used for evaluations of others' emotional experiences. Happiness is the most uniformly displayed facial expression [3]. Generally, smiles are associated with happiness or positive feelings. Discriminating smiles has many potential implications in human computing and interaction research, including marketing, advertising and social deficit intervention programs [14]. In collaborative virtual environments, avatar realism increases copresence while decreasing self-disclosure, but there are behavioural and form limitations to current Virtual Reality avatars [2], thus the question whether avatar smiles look real or posed to human observers remains an open question.

In previous work, Frank et al. [8] showed smilers' paired and single videos to observers and reported that observers were able to recognize real smiles at a rate of 74% for paired smiles and 56% for single smiles. Gosselin et al. [9] achieved a single rate of 53-57%. Del Guidice and Colle [5] used trained actors for real and acted smiles, for a single rate of 66%. Gunnery and Ruben [10] reported that female observers are more accurate than male observers in recognition of real smiles, and that the recognition rate is higher for videos compared to image observations. In our study we are interested in observers' verbal responses to viewed smilers' videos, either paired or singly. We use the expression "smiler" to indicate the person in the video performing a real or posed smile, whereas the "observer" is the person watching the video. When the same smiler was viewed by observers in both real and posed smile forms, we report this as "paired", otherwise we use the term "single". It has been suggested that "most fundamental application of affective computing would be Human-Computer Interaction (HCI) in which the computer is able to detect and track the user's affective states, and make corresponding feedback" [24]. Our work is complementary, in assuring human recognition of computer similated affecting states.

2 MATERIALS AND METHODS

Smilers' videos were collected from benchmark databases in the literature. When the smiles were elicited by showing a sequence of funny or otherwise pleasant video clips, we call them real smiles. When participants were asked to perform or instructed to display a pleasant smile, we call them posed smiles.

The collected smilers' videos were processed using oval masks in our early experiments to keep the face portion only, and presented to the observer in an order balanced way to avoid any order effects. All observers had normal or corrected to normal vision and provided written consent prior to participation. Approval from our University's Human Research Ethics Committee was received. The verbal responses were recorded from

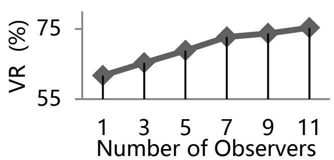

Figure 4: Obs. voting results in E1.

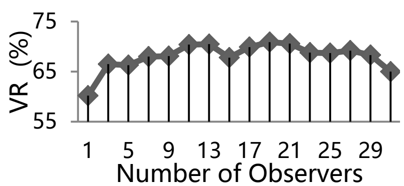

Figure 5: Obs. voting results in E2.

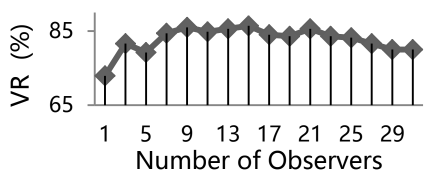

Figure 6: Obs. voting results in E3.

Table 4: Obs. VR (%) in E3

O1	*O2*	*O3*	*O4*	*O5*	*O6*
100	60	80	80	100	80
O7	*O8*	*O9*	*O10*	*O11*	*O12*
60	80	60	60	60	80
O13	*O14*	*O15*	*O16*	*O17*	*O18*
80	60	60	60	80	100
O19	*O20*	*O21*	*O22*	*O23*	*O24*
60	100	60	80	80	80
O25	*O26*	*O27*	*O28*	*O29*	*O30*
60	60	60	80	60	80
O31	***Avg.***				
60	**72.9**				

observers in four experiments watching smilers on the screen (Table 1). In Experiment I (E1), twenty-four smilers' videos (9 real, 15 posed) were collected [6, 19, 21, 23] and presented to the observers. Inspired by E1 [11, 12], E2 used a balanced set of videos (10 real and 10 posed) collected from 4 benchmark databases [7, 17, 19, 23]. We increased the number of observers from 11 in E1 to 31 in E2, but still find consistent results [13]. In E3 we used 31 new observers for 5 pairs of smilers' videos, from the UvA-NEMO database [7]. E4 used 15 observers on 14 videos: 6 human smile videos from E2 and E3 and 8 avatar smile videos [16]. See Figs. 1, 2, 3 respectively for sample frames formatted alike of real smiles, posed smiles, and avatar smiles.

3 RESULTS AND DISCUSSION

Understanding human smiles can effect and facilitate the recognition of the smiler's affective state. Avatar realism is known to affect observer behavior [1] but smile realism has not previously been examined. Each observer is not likely to be equally accurate in understanding a human smile's nature or discriminating real and posed smiles. So how many observers do we need to discriminate real and posed smiles, to maximize accuracy? To address this question, we consider all of our experiments. In E1 where 24 smiles were shown to 11 observers, their verbal response rates (VR) are depicted in Table 2 where 'O' denotes the observer. We can see that observers are on average 61.7% correct, which is somewhat above the chance value of 50%. We can improve results by combining the results of Observers, by voting, returning the decision when more than 50% of the observers detect smiles correctly. The combined results do discriminate real and posed smiles better, and improve the performance as shown in Fig. 4. The voting results are generated with the all of possible combination of observers, so the column for '3' is the average of all 165 ways to choose 3 out of 11 observers, getting their voting result, and then averaging. It is clear from Fig. 4 that votes from largest number of observers (11 observers in this case) provide the best correct verbal response rate (75.3%). So we performed a further experiment E2 where another 31 observers were shown another set of 20 smiles, again to discriminate between real and posed smiles. The results are depicted in Table 3. We can see from Table 3 that from this group of 31, 6 observers have lower than chance, 3 observers have equal to chance, and the others are above chance in discriminating real and posed smiles. On average they are 60.2% correct, this is very similar to the results we found for E1. Then, we used the voting process as in E1 to find the numbers of observers to get the best overall verbal response rate. The results are explored in Fig. 5.

We can see from Fig. 5 that VR is increased when the number of observers is increased from 1 to 11, then suddenly decreased; increased again at 19, and then gradually decreased while the number of observers is still increasing. The best verbal response rate is found in between 70.4% to 70.9% when the numbers of observers are 11, 13, 19, or 21. The curve has an overall peak between 11 and 21, but with limited differences between the top values in this range. The voting result is lower than in E1, probably due to differences in the smile videos used. We investigated this result further by implementing another experiment (E3) in paired smiles (where the same smiler was shown in both real and posed smiles). Another 31 observers were recruited to take part in this experiment, using yet another set of smile videos. The results are depicted in Table 4.

Each observer scores above chance in discriminating real and posed smiles, which is plausibly from the opportunity to compare the real and posed smiles by the same person. This improvement in the results (from 60.2% to 72.9% in our case) was also shown in Frank's [8] results. Voting again improves the overall

Table 5: Avg. VR of male and female

	E1	*E2*	*E3*
Male	59.0%	56.4%	71.3%
Female	65.0%	65.4%	75.7%
p (t-test)	0.09	**0.03**	0.16

Table 6: Votes (%) from male (M) and female (F)

No. of Obs.	*E1*		*E2*		*E3*	
	M	F	M	F	M	F
1	59.0	65.0	56.4	65.4	71.3	75.7
3	64.6	68.3	58.9	70.3	77.4	84.3
5	68.8	70.8	60.0	72.0	79.4	88.5
7	–		60.8	72.8	80.5	91.4
9			61.2	73.3	81.0	94.0
11			61.3	73.7	80.7	96.9
13			60.9	75.0	80	100
15			59.7	80.7	80	–
17			57.5	–	80	

Table 7: Realness of Avatar smiles: No. of Obs. to select avatar as ….

Smiles	Real	Fake	% Real
A1	2	13	13%
A2	2	13	13%
A3	5	10	33%
A4	2	13	13%
A5	1	14	7%
A6	***8***	***7***	***53%***
A7	*5*	*10*	*33%*
A8	*6*	*9*	*40%*

performance as explored in Fig. 6. We can see from Fig. 6 that the best VR is found for 9 (86.0%), 13 (85.6%), 15 (86.4%), and 21 (85.6%) observers. In all other cases, observers' VR gradually decrease as we move away from this region of the figure. We can conclude that 9 to 21 (or even 11 to 15) observers are enough to find the best accuracy to discriminate real and posed smiles, based on 3 sets of observers of 3 different sets of smiles.

Others have found female observers are more accurate in this context [5, 6]. Our results are reported in Table 5, and we can see that female observers are more accurate in discriminating between real and posed smiles than male observers in our experiments also. According to the two-sample one-tailed t-test we performed, these results are significant in the case of single smiles in E2, but not significant in the case of paired smiles. The higher accuracy of female observers suggests that females are more sensitive to and better understand facial cues than male observers. This effect was also found from their voting results as depicted in Table 6. It is clear from Table 6 that female observers are more accurate in discriminating between real and posed smiles in any voting combination. In E2, the best verbal rate is 61.3% when the number of male observers is 11 and 80.7% when the female observer is 15. In E3, these numbers are 11 for male and 13 for female observers. Thus a range for number of observers between 11 and 15 to get a good accuracy remains plausible. Conclusively, our result on the number of observers needed has potential applicability in social/information settings since we could chose the number of observers to correctly judge a true facial expression, in a principled fashion based on the results of our investigations as reported here. In future, we would aim to analyze observers' various physiological responses in such experiments to justify the compatibility of observers' verbal response rate with their own physiological response.

The limitations of our work are the relatively low level of accuracy of all human beings in recognizing the difference between real and posed smiles from single instances – we meet others' smiles singly, particularly when we meet new people. We have found only 1 of 8 avatars from loom.ai [16] (see also [22]) rated as real, and even then only 53% of observers rated that avatar smile as real, as seen in Table 7. It is clear that 4 of 8 avatars were very clearly rated as posed, with 3 of 8 rated as posed but with less certainty. The 53% for the single genuine avatar is a low figure, compared with the results in Table 6 for E1 and E2 on single smiles. This indicates that the best avatar smile is still less plausible as a real smile for our participants. For E4, we had chosen at random 3 real smile videos from E2 and 3 posed from E3. The results from E4 participants on these human smile videos were compatible with the results in E2 and E3, indicating that our observers in E4 behaved normally, consistently with E2 and E3.

It has been suggested that human-robot interaction is likely to be facilitated by human-like facial expressions [15] – but will we feel the robots are expressing real emotions? If not, our emotional reactions are likely to hinder appropriate interactions. Thus, for human beings observing avatar smiles that do not "feel" real, this will most likely impact on achieving the objectives for which the avatar was employed, e.g. [18]. With increasing use of avatars in service delivery and customer relations by business and government, evaluating the perceived quality of avatars will be a significant area of human computer interaction research into the future. Out future work will involve creation of avatars from our benchmark smile videos using a range of avatar creation tools to investigate: if real smiles and posed smiles remain consistent with any of the tools; use of other emotions with avatars such as anger [4]; observer training by overt feedback from their eye gaze [25]; and investigating interactions of genuineness of avatar smiles with other aspects of an avatar's visual style [20].

REFERENCES

[1] Akshay Asthana, Conrad Sanderson, Tamás D. Gedeon, and Roland Goecke. 2009. Learning-based face synthesis for pose-robust recognition from single image. In *Proc. Brit. Mach. Vis. Conf.* (BMVC '09). British Machine Vision Association and Society for Pattern Recognition, London, U.K. 1-10. DOI: https://doi.org/10.5244/C.23.31

[2] Jeremy N. Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. 2006. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments* 15, 4, 359-372.

[3] M. G. Calvo, A. Gutiérrez-García and D. Líbano. 2016. What makes a smiling face look happy? Visual saliency, distinctiveness, and affect. *Psychological Research*, 1–14.

[4] Lu Chen, Tom Gedeon, Md Zakir Hossain, and Sabrina Caldwell, 2017, Are you really angry? Detecting emotion veracity as a proposed tool for interaction. In *Proceedings of the 29th Australian Conference on Human-Computer Interaction* (OzCHI '17), Brisbane, QLD, Australia, November 2017, 5 pages. DOI: https://doi.org/0000001. 0000001

[5] M. Del Giudice and L. Colle. 2007. Differences between children and adults in the recognition of enjoyment smiles. *Developmental Psychology* 43 (2007), 796–803.

[6] Abhinav Dhall, Roland Goecke, Jyoti Joshi, Karan Sikka, and Tom Gedeon. 2014. Emotion Recognition In The Wild Challenge 2014: Baseline, Data and Protocol. In *Proceedings of the 16th International Conference on Multimodal Interaction* (ICMI '14). ACM, New York, NY, USA, 461-466. DOI: https://doi.org/10.1145/2663204. 2666275

[7] Hamdi Dibeklioğlu, Albert Ali Salah, and Theo Gevers. 2012. Are you really smiling at me? Spontaneous versus posed enjoyment smiles. In Proceedings of the 12th European conference on Computer Vision – Volume Part III (ECCV'12), Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid (Eds.), Vol. Part III. Springer-Verlag, Berlin, Heidelberg, 525-538. DOI=http://dx.doi.org/10.1007/978-3-642-33712-3_38

[8] M. G. Frank, P. Ekman and W. V. Friesen. 1993. Behavioral Markers and Recognizability of the Smile of Enjoyment. *Journal of Personality and Social Psychology* 64, 1 (1993), 83-93. DOI: https://doi.org/ 10.1037/0022-3514.64.1.83

[9] P. Gosselin, M. Perron, M. Legault and P. Campanella. 2002b. Children's and Adults' Knowledge of the Distinction between Enjoyment and Nonenjoyment Smiles. *Journal of Nonverbal Behavior* 26, 83-108.

[10] S. D. Gunnery and M. A. Ruben. 2016. Perceptions of Duchenne and non-Duchenne smiles: A meta-analysis. *Cognition and Emotion* 30, 3 (2016), 501-515.

[11] M. Z. Hossain, T. Gedeon, R. Sankaranarayana, D. Apthorp and A. Dawel. 2016. Pupillary Responses of Asian Observers in Discriminating Real from Fake Smiles: a Preliminary Study. *10th International Conference on Methods and Techniques in Behavioral Research* (Measuring Behavior'16), Dublin, Ireland, 170-176. Retrieved from http://users.cecs.anu.edu.au/~u5710140/Paper/Pupillary.pdf

[12] Md Zakir Hossain and Tom Gedeon. Observer's Galvanic Skin Response for Discriminating Real from Fake Smiles. *Australian Conference on Information Systems* (ACIS '16), Wollongong, 1-8. Retrieved from http://ro.uow.edu.au/ acis2016/papers/1/33

[13] Md Zakir Hossain and Tom Gedeon. 2017. Classifying Posed and Real Smiles from Observers' Peripheral Physiology. *11th International Conference on Pervasive Computing Technologies for Healthcare* (PervasiveHealth '17), EAI Conference Series, Barcelona, Spain, 1-4. Retrieved from http://bit.ly/2gkllCZ

[14] L. Johnston, L. Miles, and C. N. Macrae. 2010. Why are you smiling at me? Social functions of enjoyment and non-enjoyment smiles. *British Journal of Social Psychology* 49, 1 (2010), 107–127.

[15] Christine Laetitia Lisetti, Sarah M. Brown, Kaye Alvarez, and Andreas H. Marpaung. 2004. A social informatics approach to human-robot interaction with a service social robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34, 2 (2004), 195-209. DOI: https://doi.org/10.1109/TSMCC. 2004.826278

[16] Loom.ai. 2017. Personalized 3D Avatar API. Retrieved from http://www.loomai.com

[17] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. 2010. The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. *Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, 94-101.

[18] Már Másson Maack. 2017. Meet Nadia, the scarily 'human' chatbot who can read your emotions. Retrieved from http://bit.ly/2yyZ9jb

[19] M. Pantic, M. Valstar, R. Rademaker, and L. Maat. 2015. Web-Based Database for Facial Expression Analysis. *Int'l Conf. Multimedia and Expo* (ICME '05), Amsterdam, Netherlands, 317-321.

[20] Jisu Park. 2017. The Effect of Character's Visual Style on Viewer's Empathy and Sympathy Responses. Master's thesis. Purdue University, West Lafayette, Indiana. Retrieved from http://bit.ly/2yEYj3F

[21] S. Petridisa, B. Martineza, and M. Pantica. 2013. The MAHNOB Laughter database. *Journal of Image and Vision Computing* 31, 2 (2013), 186-202.

[22] Mike Seymour, Chris Evans, and Kim Libreri. 2017. Meet Mike: epic avatars. In ACM SIGGRAPH 2017 VR Village (SIGGRAPH '17). ACM, New York, NY, USA, Article 12, 2 pages. DOI: https://doi-org/10.1145/3089269. 3089276

[23] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. 2012. A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Trans. Affect. Comput.* 3, 1 (January 2012), 42-55. DOI: http://dx.doi.org/10.1109/T-AFFC.2011.25

[24] Zhihong Zeng, Jilin Tu, Ming Liu, Tong Zhang, Nicholas Rizzolo, Zhenqiu Zhang, Thomas S. Huang, Dan Roth, and Stephen Levinson. 2004. Bimodal HCI-related affect recognition. In *Proceedings of the 6th international conference on Multimodal interfaces* (ICMI '04). ACM, New York, NY, USA, 137-143. DOI: http://dx.doi.org/10.1145/1027933.1027958

[25] Dingyun Zhu, Tom Gedeon, and Ken Taylor. 2011. "Moving to the centre": A gaze-driven remote camera control for teleoperation. *Interact. Comput.* 23, 1 (January 2011), 85-95. DOI: http://dx.doi.org/10.1016/j.intcom.2010.10.003