Are paired or single stimuli better to recognize genuine and posed smiles from observers' galvanic skin response?

JESSICA SHARMIN RAHMAN, The Australian National University

MD ZAKIR HOSSAIN, The Australian National University

TOM GEDEON, The Australian National University

Smile recognition plays a vital role in human-human and human-computer interactions. This paper demonstrates a system to recognize the genuine and posed smiles by sensing observers' galvanic skin response (GSR), while watching sets of images and videos. The smiles were shown either in 'paired' or in 'single' forms. Here, 'paired' means that the same smiler was seen in both genuine and posed smile forms, otherwise the condition is referred to as 'single'. The GSR signals were recorded and processed, and several time-domain and frequency-domain features were extracted from the processed GSR signals. Classification accuracies were found to be as high as 93.6% and 91.4% from paired and single conditions respectively. In comparison, observers were verbally 59.8% and 56.2% correct. Our results demonstrate that human subconscious responses (i.e. GSR signals) is better than their own verbal response, where the paired condition is slightly better than the single condition.

CCS Concepts: • **Human-centered computing** → **Interaction techniques**.

Additional Key Words and Phrases: Galvanic Skin Response, Emotion Recognition, Smile Detection, Paired Stimuli, Classification

ACM Reference Format:

Jessica Sharmin Rahman, Md Zakir Hossain, and Tom Gedeon. 2020. Are paired or single stimuli better to recognize genuine and posed smiles from observers' galvanic skin response?. In *OzCHI '20: 32ND AUSTRALIAN CONFERENCE ON HUMAN-COMPUTER INTERACTION*, December 03–05, 2020, Online. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

A smile is a facial expression that contributes to reducing stress, increasing productivity and facilitating happiness in our lives. It is a vital non-verbal behaviour universally recognized among human beings. In the past, smiles have been categorized as Duchenne and non-Duchenne smiles, the former being an indication of a genuine smile and the latter being posed. Distinguishing between genuine and posed smile is important in order to understand distinct social cues. Studies have shown that people have correlated Duchenne smiles with positivity and preferred people displaying this smile in social settings [11]. Over the years, many researchers in the field of emotion recognition have aimed to understand the difference between genuine and posed smiles. Littlewort-Ford et al. [10] used a support vector machine (SVM) to classify Duchenne and non-Duchenne smiles and reached upto 87% accuracy. Wu et al. [16] applied discriminative local spatial-temporal descriptors in facial expression images and reached an accuracy of 91.4% in detecting spontaneous and posed smiles. Despite extensive research, it has become increasingly difficult to distinguish between different smiles using state-of-the-art methods, specially with the emergence of technologies such as Deepfakes, face morphing and so on [14].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2020 Association for Computing Machinery.

Manuscript submitted to ACM

Physiological signals have shown tremendous results in detecting different emotions among people. Perusquia-Hernandez et al. [12] was able to differentiate between spontaneous and posed smiles using features from facial EMG. Signals such as pupillary response and skin conductance also shown to accurately predict upto 97.8% of the time in identifying genuine and posed smiles [6]. Although observers' physiological signals have shown promising results in recognizing genuine and posed smiles, it is not clear if they are better in recognizing these different smiles in a single or paired condition. Here, paired condition means observers were shown both genuine and posed smiles from same smiler and single condition means each smile was different with respect to smiler. In the paired condition, one stimuli was genuine, while the other was posed.

We investigate the changes in observers' galvanic skin response (GSR) signals in distinguishing between genuine and posed smiles in single and paired conditions. GSR signals have been used extensively in recognizing emotions as they are sensitive to different emotional states [9]. GSR signals were recorded from observers while they watched a series of stimuli containing genuine and posed smiles in different conditions. Three machine learning techniques were applied to classify between genuine and posed smiles using the GSR signals. The rest of the sections of this paper are organized as follows: Section 2 explains the experiment design, materials and techniques used for the experiment. Section 3 describes the results and discusses some implications from the results. Finally in section 4, the paper is concluded by highlighting some limitations and future work.

2 MATERIALS AND METHODS

2.1 Observers

A total of 25 participants (called observers subsequently) volunteered to participate in this study, being 13 male and 12 female. Their mean age was 21.3 ± 2.6 . Observers' GSR was recorded using the Empatica E4 watch-format device [4]. The signals were recorded at a sampling rate of 4Hz. The study was approved by the Human Research Ethics Committee of the Australian National University (ANU). The observers were asked to sign a written consent form before the experiment, followed by observers providing some demographic information. Then they were presented with a series of stimuli and asked some questions. All experiments were conducted in the same location in order to minimize any environmental effects. Verbal response from the observers were collected using an interactive website.

2.2 Stimuli

The stimuli used in this study were selected from UvA-NEMO database [3] which consists of stimuli showing genuine and posed smiles. There were a total of four conditions in which participants viewed the stimuli. They are: Single Image (SI), Paired Image (PI), Single Video (SV) and Paired Video (PV). In SI and SV conditions, observers were presented with a single image or video of smiles and then they were asked if they thought it was genuine or posed. In PI and PV conditions, observers watched images and videos in pairs and they were asked which one of them they thought was genuine. The stimuli that were presented in pairs were from the same smiler. Figure 1 shows sample frames from the database in both single and paired conditions.

For both SI and SV conditions 10 stimuli were used, out of which five were genuine and five were posed. For the PI and PV conditions, 20 stimuli were used (10 genuine and 10 posed). Altogether there are 20 stimuli in the single condition and 40 in the paired condition. All observers (i.e. participants) watched all stimuli. They were order balanced in order to remove sequence bias. The image stimuli were shown for five seconds each while the video stimuli varied from two to seven seconds in length. From the 25 observers' data, four were discarded due to poor or missing GSR

Fig. 1. Sample Frames From UvA-NEMO Database: (a), (b) = paired condition, (c), (d) = single condition

Figure 1 description Four sample frames (three female and one male) shown from UvA NEMO database. Frame (a) and (c) corresponds to real smile, and (b) and (d) corresponds to posed smile

recordings. This sometimes occurred due to the poor connection quality or low charge of the device. Therefore, 21 observers' data were used for further analysis.

2.3 Data Preprocessing

GSR signals are prone to noise due to subject movements. The signal range also varies from person to person. Therefore, some pre-processing techniques need to be performed in order to reduce the noise in the signals. We used a min-max normalization method to normalize the signals within the 0-1 range. This is done to remove any between participant differences. In order to remove noise from the signals, we used the median smoothing filter which is frequently used to filter physiological signals [8]. Then we segmented the signals according to the length of stimuli. For example, the images used in the experiment were shown for five seconds. Thus, we extracted the five seconds' data corresponding to the respective image stimulus from the whole data recorded for a single participant in an excel file, and we call these a 'segment'.

Raw physiological signals collected from humans often cannot provide much useful information. It also often results in a large dataset which becomes computationally expensive to analyze. Thus, we extracted a number of time domain and frequency domain features from our segmented GSR signals. Based on some papers in the literature focusing on sensor based signal processing [2, 13], we extracted the features shown in Table 1.

Table 1. Features Extracted from GSR Signals

Feature Type	Feature Names
Time domain features	Mean, minimum, maximum, standard deviation, variance, root mean square, summation, absolute summation, simple square integral, mean of first and second difference of normalized signals, Hjorth parameters (mobility), simple square integral, log detector
Frequency domain features	Mean, minimum, maximum of the first 16 points from Welch power spectral density

Twenty five features were initially extracted from each segment of GSR signals. We visualized the features and removed some features which could not give much insights into the data. For example, we calculated the number of

peaks for each segment. However, due to the small size of each segment, we could not notice many peaks during each segment. Similarly we discarded some features where the values were mostly zeroes or ones. We further removed some redundant features, and this resulted in a total of 16 features from each segment which are those listed in table 1.

2.4 Classification and Evaluation Measures

Selected features from the segmented GSR signals were then trained using three different classification techniques. They are, Decision Trees (DT), K-Nearest Neighbor (KNN) and Bagged Trees (BT). Cross validation was done using a leave-one-observer-out method. In this approach, one observer's data is used for testing and the others are used for training the model. This approach is then done in an iterative manner for all observers to validate the model. Training and testing was done in both single and paired conditions. The prediction accuracies were compared with both the labels provided by the database and observers' verbal response labels. In addition to accuracies, we also calculated the precision (positive predictive value), recall (true positive rate), specificity (true negative rate) and f-measure (harmonic mean of precision and recall) as our evaluation measures. The complete data analysis was done using MATLAB® R2020a software with AMD Ryzen 7 3700X 8-Core Processor with 3593 Mhz, 16.00 GB of RAM and Microsoft Windows 10 Home 64-bit operating system.

3 RESULTS AND DISCUSSION

Classification accuracy obtained from observers GSR signal features using KNN, DT and BT methods are reported in Table 2. The first three rows correspond to the accuracies using the labels provided by the database (objective ground truth). The following three rows are results obtained by comparing with observers' verbal response labels (subjective 'ground truth').

Table 2. Classification Accuracy Using GSR Signals

Condition	Label	KNN	DT	BT
Single	UvA-NEMO Label	59.04%	91.43%	90.95%
Paired	UvA-NEMO Label	68.1%	90.95%	86.07%
Paired (Subset)	UvA-NEMO Label	71.9%	93.57%	88.57%
Single	Observers' Verbal Response	48.33%	56.19%	52.14%
Paired	Observers' Verbal Response	60.23%	58.57%	62.14%
Paired (Subset)	Observers' Verbal Response	60.95%	59.76%	65.24%

We can see from table 2 that the highest accuracy is achieved by DT in a paired condition (93.6%). Initially we trained the models using our entire dataset and found that the accuracy is higher in the single condition (91.4%) in comparison to paired condition (90.9%). However, we suspected that this could be due to the imbalance in the dataset. As explained in section 2, we used a total of 20 stimuli in single condition and 40 in the paired condition. Therefore, we randomly chose a subset of 20 stimuli (10 genuine and 10 posed) from the paired condition to train the classifiers. Results from table 2 demonstrate that the accuracy is higher than the single condition for all three classification methods. Thus, our results indicate that the paired conditions is better than the single condition for observers to distinguish between genuine and posed smiles.

Classification results of the other evaluation measures were also higher in the paired condition in comparison to the single condition. Precision, recall, specificity and F-measure scores of the DT model in the paired condition were 93.8%,

93.3%, 93.8% and 93.6% respectively. The same evaluation measures in the single condition were 92.2%, 90.5%, 92.4% and 91.3%. Classification accuracy of a model often fails to provide complete insights on the predictive power of the model [15]. The high scores of our model in all evaluation measures confirms our result and suggests that stimuli shown in pairs are more beneficial than showing as a single stimulus.

In order to understand how well the observers' GSR signals are distinguishing between genuine and posed smiles, we performed a timeline analysis of the pre-processed GSR signal averages in both single and paired condition. We reshaped all the signals to have the initial value 0.5. We chose this initial value in order to clearly observe the increasing or decreasing trend of the GSR signals. We also selected values for the first five seconds of each stimuli as this is the length of most stimuli used in the experiment. The results are shown in Figure 2. Red shaded area displays GSR signals when observers are watching genuine smiles and blue shaded areas correspond to watching posed smiles.

Fig. 2. Timeline Analysis of GSR signals

Figure 2 description Two shaded line graphs showing the average GSR values of all participants watching real and posed stimuli. X axis shows the time in seconds, starting from 0 to 5 seconds. Y axis shows the GSR values, starting from 0 to 0.51. Detailed explanation of the figure is given in the discussion section.

It can be seen that observers' GSR signals are displaying some differences in range from the first second of watching a single genuine and posed stimuli. The difference becomes larger after 3.5 seconds, which suggests that observers' response becomes clearer after watching the stimuli for a few seconds. In the case of paired stimuli, the GSR signals are quite similar for both genuine and posed smiles in the first second, but it starts diverging from each other quickly after that. And after three seconds, both signals show larger difference in their values. In both cases, observers' GSR signals are able to distinguish between genuine and posed stimuli after three seconds. This finding is useful in designing future studies in this area, where shorter length stimuli should not be used. It is also noticed that GSR signals are higher when they are presented posed stimuli, which shows that their subconscious reactions to posed stimuli is stronger compared to genuine stimuli as found in another study considering pupillary responses [5]. We further verified the significance level of GSR signals in both single and paired conditions using a Kolmogorov–Smirnov (K-S) test. The results showed statistical significance ($p < 0.05$) between genuine and posed smile signals for both conditions.

Although observers' GSR response aligned very well with the objective ground truth information, it does not align well when compared with their verbal response. We can further see from table 2 that the highest accuracy obtained

using observers' verbal response was 65.2% with the BT method, which is above chance only to a limited extent. In the single conditions, the accuracy was above chance using the DT method (56.2%), and lower than chance (48.3%) with the KNN method. This indicates that observers perform poorly in discriminating genuine and posed smiles when they are watching only one stimuli at a time. However, when shown in pairs, they perform slightly better. Overall, our analysis shows that observers' subconscious GSR signals are better in recognising genuine and posed smiles when compared to their own verbal responses. In addition, our studies revealed usage of paired stimuli to be more effective than using single stimuli. This phenomena has not been explored before, even though there have been studies that showed significant differences in emotional and physiological response watching genuine and posed stimuli [1, 5].

4 CONCLUSION

This paper presents a preliminary study to collect and analyze observers' GSR signals while they watch stimuli containing genuine and posed smiles in single and paired conditions. Experimental results show that observers' GSR signals are 91.4% accurate in differentiating between genuine and posed smiles when they watch a single stimuli, and 93.6% accurate when they watch stimuli in pairs. This study also reveals that observers' verbal responses perform poorly compared to their physiological responses, as their correct response rate was only 56.2% and 59.8% accurate in single and paired conditions respectively. The effect was also evident in that the verbal response labels were harder to classify by any of the methods, with a maximum of 65.2% achieved. It is clear that the paired stimuli helped all techniques: the AI tool on ground truth (UvA-NEMO) labels was helped the least, followed by the verbal response, while the AI tools on the verbal response labels were improved by 9.8% on average. This suggests that the improvement is due to humans finding comparisons easier than absolute judgements.

There are certain limitations of this work. Due to the small number of participants, we did not have enough sample data to build a more robust model. Future work will include collecting more data and other physiological signals such as blood volume pulse, electroencephalogram, skin temperature etc. We will also collect detailed verbal response from observers in order to understand what difficulties they faced to identify the different smile types. Further, we will consider more robust feature selection methods, such as CCA Network [7], to find reliable features, and will compare our results with computer vision approaches, such as RealSmileNet [17], in terms of both computational complexity and reliability. Nevertheless, these preliminary findings provide evidence of the usefulness of observers' subconscious response in differentiating between genuine and posed emotions.

ACKNOWLEDGMENTS

The authors would like to thank all the participants of this study.

REFERENCES

- [1] Claudio Aracena, Sebastián Basterrech, Václav Snáel, and Juan Velásquez. 2015. Neural networks for emotion recognition based on eye tracking data. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2632–2637.
- [2] Rubana H Chowdhury, Mamun BI Reaz, Mohd Alauddin Bin Mohd Ali, Ashrif AA Bakar, Kalaivani Chellappan, and Tae G Chang. 2013. Surface electromyography signal processing and classification techniques. *Sensors* 13, 9 (2013), 12431–12466.
- [3] Hamdi Dibeklioglu, Albert Ali Salah, and Theo Gevers. 2012. Are you really smiling at me? spontaneous versus posed enjoyment smiles. In *European Conference on Computer Vision*. Springer, 525–538.
- [4] E4 [n.d.]. E4 Wristband from Empatica. <https://www.empatica.com/research/e4/>
- [5] MZ Hossain, Tom Gedeon, R Sankaranarayana, Deborah Apthorp, and Amy Dawel. 2016. Pupillary responses of Asian observers in discriminating real from fake smiles: A preliminary study. In *Measuring Behavior*. 170–176.
- [6] Md Zakir Hossain and Tom Gedeon. 2019. Observers' physiological measures in response to videos can be used to detect genuine smiles. *International Journal of Human-Computer Studies* 122 (2019), 232–241.

- [7] Md Zakir Hossain, Md Monirul Kabir, and Md Shahjahan. 2016. A robust feature selection system with Colin's CCA network. *Neurocomputing* 173 (2016), 855–863.
- [8] S Jerritta, M Murugappan, R Nagarajan, and Khairunizam Wan. 2011. Physiological signals based human emotion recognition: a review. In *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*. IEEE, 410–415.
- [9] Jonghwa Kim and Elisabeth André. 2008. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence* 30, 12 (2008), 2067–2083.
- [10] Gwen Littlewort-Ford, Marian Stewart Bartlett, and Javier R Movellan. 2001. Are your eyes smiling? detecting genuine smiles with support vector machines and gabor wavelets. In *In Proceedings of the 8th Joint Symposium on Neural Computation*.
- [11] Marc Mehu, Anthony C Little, and Robin IM Dunbar. 2007. Duchenne smiles and the perception of generosity and sociability in faces. *Journal of Evolutionary Psychology* 5, 1 (2007), 183–196.
- [12] Monica Perusquía-Hernández, Mazakasu Hirokawa, and Kenji Suzuki. 2017. Spontaneous and posed smile recognition based on spatial and temporal patterns of facial EMG. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 537–541.
- [13] Jessica Sharmin Rahman, Tom Gedeon, Sabrina Caldwell, Richard Jones, Md Zakir Hossain, and Xuanying Zhu. 2019. Melodious Micro-frissons: Detecting Music Genres From Skin Response. In *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.
- [14] Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales, and Javier Ortega-Garcia. 2020. Deepfakes and beyond: A survey of face manipulation and fake detection. *arXiv preprint arXiv:2001.00179* (2020).
- [15] Francisco J Valverde-Albacete and Carmen Peláez-Moreno. 2014. 100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox. *PloS one* 9, 1 (2014), e84217.
- [16] Pingping Wu, Hong Liu, and Xuewu Zhang. 2014. Spontaneous versus posed smile recognition using discriminative local spatial-temporal descriptors. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1240–1244.
- [17] Yan Yang, Md Zakir Hossain, Tom Gedeon, and Shafin Rahman. 2020. RealSmileNet: A Deep End-To-End Network for Spontaneous and Posed Smile Recognition. *arXiv preprint arXiv:2010.03203* (2020).