Classifying Posed and Real Smiles from Observers' Peripheral Physiology

Md Zakir Hossain
The Australian National University
Canberra 2601, Australia
zakir.hossain@anu.edu.au

Tom Gedeon
The Australian National University
Canberra 2601, Australia
tom@cs.anu.edu.au

Abstract
Smiles are important signals in face-to-face communication that provides impressions / feelings to observers. For example, a speaker can be motivated from audience smiles. People can smile from feeling or by acting or posing the smile. We used observers' physiological signals such as PR (Pupillary Response), BVP (Blood Volume Pulse), and GSR (Galvanic Skin Response) to classify smilers' real (elicited) and posed (asked to act) smiles. Twenty smile videos were collected from benchmark datasets and shown to 24 observers while asking them to make choices, and recording their physiological signals. A leave-one-video-out process was used to measure classification accuracies, and was 93.7% accurate for PR features.

Author Keywords
Observers; Posed and Real Smiles; Physiological Signals; Classification; Affective Computing

PervasiveHealth '17, May 23–26, 2017, Barcelona, Spain

ACM ISBN 978-1-4503-6363-1/17/05…$15.00
https://doi.org/10.1145/3154862.3154893

ACM Classification Keywords
H.1.2 Human information processing;
H.5.5. Signal analysis, synthesis, and processing.

Introduction
Life provides many reasons to smile that most often indicate pleasure, appreciation, happiness, or satisfaction. A smiling face evokes positive feelings and can elicit return smiles [1]. It is also known that smiles in yearbook photos are positively correlated with preventative health interactions with the medical profession [2]. Thus it is plausible that seeing real smiles is more beneficial for health than posed smiles. Our work provides a tool to determine if the observer of a smile feels the smile to be real, as that is reflected in the observer's physiological signals. Our research serves as a platform for further work in this area.

When smilers were asked or instructed to display a smile, we consider these smiles to be *posed smiles*. When a smile is elicited by showing funny or otherwise pleasant video clips, we consider them to be *real smiles*. Creating a system that can classify posed and real smiles could be applicable in many situations, such as tutoring systems, video conferencing, customer service quality evaluation, depression monitoring, for truthfulness during questioning, and so on.

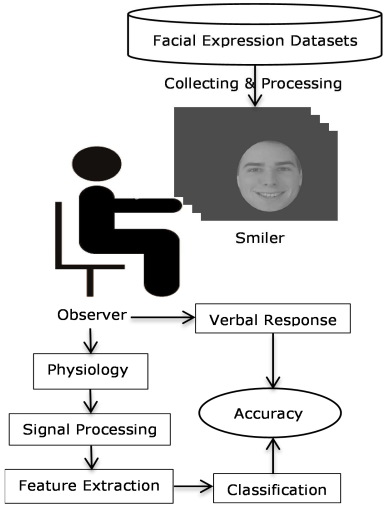


Figure 1: The framework of the research methodology.

In the past, researchers focused on analyzing smilers' faces directly and/or observers' verbal responses to classify posed and real smiles. Frank et al. [3] considered observers' verbal responses to recognize real smiles. Hamdi et al. [4] used a computer vision technique on the smilers' facial features. Hoque et al. [5] performed two experiments to classify 'delighted' and 'frustrated' smiles from smilers' facial features and observers' verbal responses. Although smiles are one of the easiest, voluntarily, and frequently performed facial expressions [4], it is not an easy task to distinguish real and posed smiles from observers' verbal responses [3]. It is possible to recognize elicited emotion from physiological signals [6]. We analyzed observers' three physiological signals – pupillary response (PR), galvanic skin response (GSR), and blood volume pulse (BVP) – to classify real and posed smiles.

A person's pupillary responses can change for many reasons, including memory load, stress, pain, watching videos, face to face interactions etc., and would offer a good method for classifying real and posed smiles [7]. GSR is an automatic reaction that measures electrical changes of human skin and is considered one of the strongest signals in emotion detection [8]. BVP is another vital physiological signal that measures blood volume changes using infrared light through the skin. We hypothesized that the involuntary nature of physiological signals may provide an improvement over observers' low classification rate for real versus posed smiles in their verbal responses. We know that feature extraction methods play a crucial role in physiological signal processing for emotion recognition [6]. We extracted six time domain statistical features from each peripheral physiological signal in each observation.

RESEARCH METHODOLOGY

The working procedure is illustrated in Figure 1.

Smilers' Videos

Twenty videos were collected from four databases (five from each), namely UvA-NEMO [9], MAHNOB [10], MMI [11], and CK+ [12]. Real smiles were collected from NEMO and MAHNOB databases where participants were induced to smile by watching a sequence of funny or otherwise pleasant video clips. Posed smiles were collected from MMI and CK+ databases where participants were asked to perform or instructed to display a smile. The videos were then processed using MATLAB R2014b to make them uniform in size, format (greyscale, mp4) and smile duration of 5 sec. Only faces of smilers are shown, with the background masked. The luminance and contrast of these videos were adjusted using MATLAB SHINE toolbox [13].

Conduct of the experiment

Twenty-four healthy, right-handed participants took part as observers in this experiment, with a mean age of 30.7±6.0 (mean±SD). They signed an informed consent form prior to their voluntary participation. The experiment was approved by the Australian National University's Human Research Ethics Committee. A 15.6" ASUS laptop and a computer mouse are peripherals for interaction between the observer and a laptop running the web-based tool showing the smile videos. The chair of the observer is moved forward or backwards to adjust the distance between the observer and eye tracker. Observers are asked to track a spot displayed in the laptop for calibrating the eye tracker and starting the experiment. Observers are instructed to limit their body movements in order to reduce undesired artefacts in the signals. The Eye Tribe

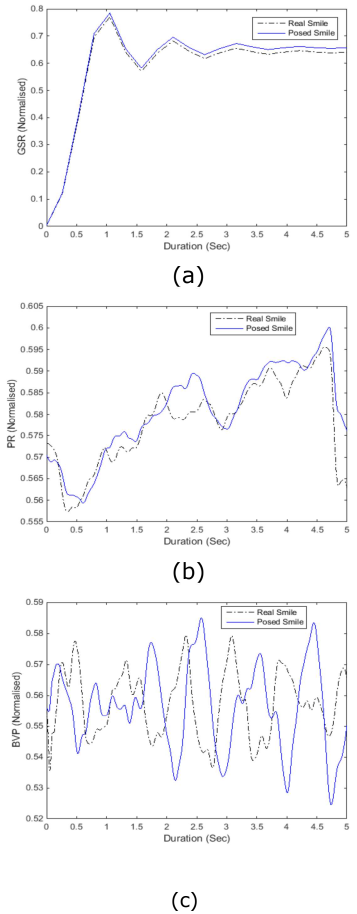


Figure 2: Average GSR (a), PR (b), and BVP (c) signal timelines for smilers' videos over all observers.

(https://theeyetribe.com/) remote eye-tracker and wireless Empatica E4 (https://www.empatica.com/) watch were used to record PR, BVP and GSR signals at a sampling rate of 60 Hz, 64 Hz and 4 Hz respectively. Videos are presented in an order balanced way followed by questions to identify smiler's real or posed smiles.

Signal Processing and Feature Extraction

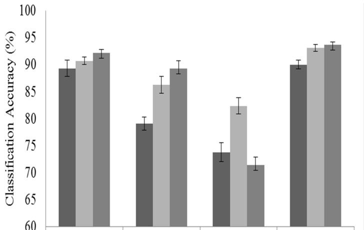
The eye blinking points of both eyes' PR signals were reconstructed using cubic spline interpolation and smoothed using 10-point Hann moving window average filter respectively [7], and then averaged them to find a single PR signal. Low-pass Butterworth filter (order = 6, cut-off frequency = 0.5) is applied to smooth the GSR and BVP signals. Then, six time domain statistical features (mean, maximum, minimum, standard deviations, and means of the absolute values of the first and second differences of the processed signals) are extracted from each video related peripheral physiological signal. These features are easy to compute and also cover the typical characteristics of the signals: range, gradient, and variation. Each signal is normalized to keep the raw and extracted features in a range between 0 and 1. Thus, there are 120 extracted features (20 videos x 6 features) for an observer and a total of 2,880 features for all 24 observers. During training, 2,736 features (24 observers x 6 features x 19 videos) are used, with 144 (24 observers x 6 features x 1 video) for testing. This leave-one-video out means that our classifiers have seen no physiological signals from any observers on that video. Our results are thus video independent.

EXPERIMENTAL RESULTS

The methods we used for smile classification are k-nearest neighbor (KNN), support vector machine (SVM), neural network (NN), and an ensemble aggregating the decision of these three classifiers. The parameters were 3 nearest neighbors, Gaussian radial basis kernel function with a scaling factor of 0.5, scaled conjugate gradient training function with 5 hidden nodes, and mean square performance function for KNN, SVM, NN, and ensemble classifiers respectively.

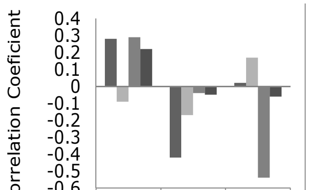
The timeline analysis of each peripheral physiological signal exposed a common trend. The skin response (Figure 2(a)) started increasing from onset till maximum at 1-1.5s and then decreased to a minimum at 1.5-2s. After that, the average GSR signal deviated in 0.55-0.6 normalized amplitudes. On the other hand, we found that pupils (Figure 2(b)) constricted from the onset and reached a minimum at 0.5-1s, after which dilation started and continued till the maximum in a zigzag fashion. It is worth noting that the trends are different according to real or posed smile observation, where a posed smile shows higher amplitudes compared to a real smile observation. The timeline analysis of BVP signal is somewhat different, where amplitude variation for a real smile observation is less compared to a posed smile observation (Figure 2(c)). Two-sample Kolmogorov-Smirnov (K-S) test show that average GSR ($p=0.0026$), PR ($p=0.0015$), and BVP ($p=0.0122$) signals differed significantly for posed smile observations compared to real smile observations.

In the analysis, we notice that (Figure 3) higher accuracies with lower standard deviations (error bars) are found from the ensemble classifier, when compared to the other classifiers, with the highest accuracy of 93.7% (±0.5) for PR features while observers are only 59.0% correct (on average), with chance being 50%. In the literature, Frank et al. [3] and Hoque et al. [5]



	NN	SVM	KNN	Ensemble
GSR	89.3	79.1	73.8	90
BVP	90.7	86.3	82.4	93.1
PR	92.2	89.3	71.4	93.7

Figure 3: Average classification accuracies for all observers over the videos.



	GSR	BVP	PR
NN	0.28	-0.42	0.02
SVM	-0.09	-0.17	0.17
KNN	0.29	-0.04	-0.54
Ensemble	0.22	-0.05	-0.06

Figure 4: Correlation between observers' verbal responses and classifiers' accuracies.

found that observers were 56.0% and 69.0% correct respectively. Analyzing the smiler's images/videos directly, they found classification accuracies of 92.9% [4], and 92.0% [5]. These results are comparable to ours, but from analysis of smiler's images/videos directly, while our results are from the observers' reactions to videos. Our technique is general and does not require specific computer code built on properties of smiles. Instead we have developed a process to analyse patterns of changes of physiological signals via machine learning, and can be applied to other facial expressions. The observers' verbal responses either negatively or only slightly correlated with classifiers' accuracies as shown in Figure 4, indicating that better verbal responses did not consistently imply better results from the physiological signals.

In conclusion, high accuracy is found at 93.7% from PR features where the same observers are themselves only 59% correct according to their verbal responses. This provides an indication that observers' peripheral physiology can be applied to reliably classifying smilers' smiles into real and posed, in a generalizable way. There are implications of our work for health beyond smiles. One is to use our techniques in other areas such as stress [14], or depression [15].

References

1. W.C. Mackey. 1976. Parameters of the smile as a social signal. *J. of Genetic Psychology*, 129(1), pp.125-130.
2. M. Cross. 2016. *Does a Smile on Picture Day Keep the Doctor Away? The Connections between Facial Expressions in Student Identification Photographs and Health Care Center Visitation*. MA Thesis, UC Irvine.
3. M.G. Frank, P. Ekman, W.V. Friesen. 1993. Behavioral markers and recognizability of the smile of enjoyment. *J Pers Soc Psychol*, 64(1), 83-93.
4. H. Dibeklioğlu, A.A. Salah, T. Gevers. 2015. Recognition of Genuine Smiles. *Trans. Multimedia*, 17(3), 279-294
5. M. E. Hoque, D. J. McDuff, R. W. Picard. 2012. Exploring Temporal Patterns in Classifying Frustrated and Delighted Smiles. *Trans. Affective Computing*, 3(3), 323 – 334.
6. P. Gong, H. T. Ma, and Y. Wang. 2016. Emotion recognition based on the multiple physiological signals, *Int. Conf. on Real-time Computing and Robotics*, 140-143.
7. 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. *Measuring Behavior*, 170-176.
8. M. Z. Hossain, T. Gedeon, R. Sankaranarayana. 2016. Observer's Galvanic Skin Response for Discriminating Real from Fake Smiles. *Australasian Conference on Information Systems*, 1-8.
9. H. Dibeklioglu, A.A. Salah T. Gevers. 2012. Are You Really Smiling at Me? Spontaneous versus Posed Enjoyment Smiles. *European Conf. on Computer Vision*, 525-538.
10. M. Soleymani, J. Lichtenauer, T. Pun, M. Pantic. 2012. A multimodal database for affect recognition and implicit tagging. *Trans. Affective Computing*, 3(1), 42 – 55.
11. M. Pantic, M. Valstar, R. Rademaker, L. Maat. 2005. Web-Based Database for Facial Expression Analysis. *International Conference on Multimedia and Expo*, 5.
12. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, 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*, CA, 94-101.
13. V. Willenbockel, J. Sadr, D. Fiset, G.O. Horne, F. Gosselin, J.W. Tanaka. 2010. controlling low-level image properties: the SHINE toolbox. *Behav Res Methods*, 42(3), 671-684.
14. N. Sharma, T.D. Gedeon. 2014. Modeling observer stress for typical real environments. *Expert Systems with Applications*, 41(5), 2231-2238.
15. S.Alghowinem, R. Goecke, M. Wagner, J. Epps, T.D. Gedeon, M. Breakspear, G. Parker. 2013. A comparative study of different classifiers for detecting depression from spontaneous speech. *IEEE ICASSP*, 8022-8026.