Using Temporal Features of Observers' Physiological Measures to Distinguish Between Genuine and Fake Smiles

Md Zakir Hossain, Tom Gedeon, *Senior Member, IEEE*, and Ramesh Sankaranarayana

Abstract—Future affective computing research could be enhanced by enabling the computer to recognise a displayer's mental state from an observer's reaction (measured by physiological signals), using this information to improve recognition algorithms, and eventually to computer systems which are more responsive to human emotions. In this paper, an observer's physiological signals are analysed to distinguish displayers' genuine from fake smiles. Overall, thirty smile videos were collected from four benchmark database and classified as showing genuine or fake smiles. Overall, forty observers viewed videos. We generally recorded four physiological signals: pupillary response (PR), electrocardiogram (ECG), galvanic skin response (GSR), and blood volume pulse (BVP). A number of temporal features were extracted after a few processing steps, and minimally correlated features between genuine and fake smiles were selected using the NCCA (canonical correlation analysis with neural network) system. Finally, classification accuracy was found to be as high as 98.8 percent from PR features using a leave-one-observer-out process. In comparison, the best current image processing technique [1] on the same video data was 95 percent correct. Observers were 59 percent (on average) to 90 percent (by voting) correct by their conscious choices. Our results demonstrate that humans can non-consciously (or emotionally) recognise the quality of smiles 4 percent better than current image processing techniques and 9 percent better than the conscious choices of groups.

Index Terms—Affective computing, genuine smile, fake smile, physiological signals, temporal features

1 Introduction

The human face conveys not only our identity, but also important information related to our emotional states or moods. Facial expression research began in 1872 by analysing observers' responses to different emotions [2]. Observers categorised six emotions in [3]. Tomkins and McCarter [4] achieved very high agreement from observer's judgements on facial expressions. Observers reliably distinguished between individuals' amusement and embarrassment in [5]. Evidence indicates that observers experience fairly specific responses to facial expressions [6]. For example, facial expressions of distress [7], anger [8], and embarrassment [9] evoke sympathy, fear, and amusement respectively. However, some emotional expressions can be complex. The facial expression generally considered as 'the smile' is not a singular category of facial behaviour [10]. This is because smiles convey not only happiness [11], but can also signify frustration, anger, depression, embarrassment, empathy, surprise, distress, polite disagreement, pain, and even more. It is the easiest facial expression to fake voluntarily [12].

Smiles also have social signal value [10], which may be relevant in many situations like public security, social interaction, human-computer interaction, and so on. As an example, a custom officer may detect that the person at passport control appears suspicious, or not. It is important to accurately differentiate between genuine and kinds of non-genuine smiles to understand the affective state differences behind the smiles.

Valstar et al. [13] used video data from expression displayers' face, head and body actions to classify smiles. Ambadar et al. [14] investigated the characteristics of smiles along with their perceived meanings, and found that perceived meanings were related to specific characteristics. Dibeklioğlu et al. [1] extracted a number of features from displayers' faces and implemented an automatic assessment technique. Cohn et al. [15] examined image data to measure the timing of face motion during smiles. Gan et al. [16] applied a two-layer deep Boltzmann machine to displayers' image data. On the other hand, self-reports of observers' judgements were considered in [10], [17], and [18] in the context of distinguishing real and posed smiles. In this paper, we will compare observers' physiological measures with their own judgements to distinguish posed, acted or social smiles, which we refer to as 'fake' or 'non-genuine' smiles, from spontaneous smiles, which we refer to as 'real' or 'genuine smiles'.

In general, recognition from video is easier for users than recognition from static images [20]. Observers may experience certain feelings [21], [22] from watching video clips or listening to music that are reflected in their physiological signals. Observers' physiological changes are also associated with their emotional states [21] and are less susceptible

- *The authors are with the Research School of Computer Science, Australian National University, Canberra, Australia.*
E-mail: zakir.hossain@anu.edu.au, {tom, ramesh}@cs.anu.edu.au.

Manuscript received 24 Nov. 2016; revised 7 Oct. 2018; accepted 20 Oct. 2018. Date of publication 25 Oct. 2018; date of current version 28 Feb. 2020.
(Corresponding author: Md Zakir Hossain.)
Recommended for acceptance by F. Schwenker.
Digital Object Identifier no. 10.1109/TAFFC.2018.2878029

1949-3045 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: CLARKSON UNIVERSITY LIBRARY. Downloaded on February 01,2022 at 16:31:54 UTC from IEEE Xplore. Restrictions apply.

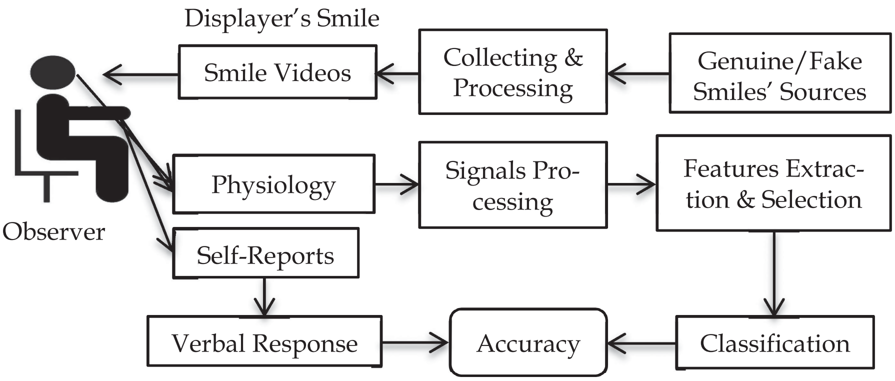

Fig. 1. Basic block diagram of the experiment.

to social masking [23] than their determinations. The physiological reactions of the body and brain change in response to different stimuli, and these reactions are possibly or partly recognized by people as they form their judgements [24], in our case about the realness or not of smiles. On the other hand, affective self-reports might be held in doubt because errors in judgements are not negligible; the observers might misrepresent or cannot always remember the different smiles seen during an experiment or want to please the experimenter [25]. People commonly misrepresent experienced emotions, whether intentionally or not, which seems to be a common occurrence in social interactions [17]. The approach of using observers' physiological signals to decode affective responses [22] to smiles is an alternative way of accessing the displayers' internal state.

In this paper, four physiological observers' signals were analysed to distinguish genuine from fake smiles: pupillary response (PR), Electrocardiogram (ECG), galvanic skin response (GSR), and blood volume pulse (BVP). The pupillary response is the measure of pupil diameter over time. Among other things, pupil diameter is influenced by light, cognition, attention, and emotion [26], [27]. The pupillary reflex has been found to vary significantly during the identification of smiles or emotions after removing luminance effects [25]. Principal Component Analysis (PCA) can be used as an alternative, to reliably separate the effect of changes in luminance from other effects [28]. ECG is also a prominent physiological signal used to distinguish different kinds of emotional states [29], [30]. As an example, unpleasant stimuli result in higher cardiac deceleration [31]. Galvanic skin response (GSR, also called skin conductance or electrodermal response) is another important physiological signal that has been found to be sensitive to emotional changes [21], [32]. Recent research has indicated that reactions to happiness and sadness can be distinguished from GSR [30]. Measurements of BVP are also commonly used in emotional state classification [33]. To the best of our knowledge, no work prior to ours has been done to distinguish genuine from fake smiles from observers' physiological signals. This paper uses feature-based classification of observers' physiological signals while watching video stimuli of smiles to distinguish their reactions to genuine from fake smiles.

We collected video clips from four benchmark databases to use as stimuli for the observers during the main study. We use the expression "displayer" to indicate the person in the video performing a genuine or fake smile, whereas the "observer" is the person watching a video. First, twenty smile video clips were selected at random from four benchmark datasets [34], [35], [36], and [37] and processed using MATLAB R2015a to convert all of these to grey scale, with each clip lasting 5s. The use of grayscale and 5s was to eliminate differences between data sources. Second, physiological signals were recorded from twenty-six observers, while watching the smile video stimuli, along with their judgments of whether each smile was genuine or fake. Third, omitting data from two observers due to technical (data recording) problems, the data from twenty-four observers were analysed. The analysis consisted of several stages: pre-processing (normalisation, smoothing, de-noising), feature extraction, feature selection, and classification. A basic blook diagram of the experiment is shown in Fig. 1. Five classifiers (neural network (NN), support vector machine (SVM), relevance vector machine (RVM), k-nearest neighbours (KNN), and Ensemble) were chosen, to classify between genuine and fake smiles by analysing the temporal features of each physiological signal. Finally, the significance levels of signals were verified using the Kolmogorov–Smirnov test (K-S test) and the permutation test. In this case, canonical correlation analysis with neural network (NCCA) was the feature selection system used to find minimally correlated features between genuine and fake smiles. The results showed that classification accuracies were much higher than chance and significantly higher than observers' self-reports. In an additional study, we used both genuine and fake smiles' stimuli from a single source [34], and collected PR and BVP signals from 16 new observers. The collected signals were analysed in similar fashion as described for the main study, obtaining consistent results.

2 Materials and Methods

2.1 Smiles Stimuli

The smiles stimuli were selected at random from four databases for the main study: UvA-NEMO (University of Amsterdam-NEMO) [34], (used videos: 002_sponteneous_smile_1, 003_sponteneous_smile_1, 018_sponteneous_smile_1, 020_sponteneous_smile_2, 069_sponteneous_smile_2), MAHNOB (Multimodal Analysis of Human Nonverbal Behaviour in Real-World Settings) [35], (videos: P2-Rec1-2009, P4-Rec1-2009, P8-Rec1-2009, P14-Rec1-2009,

Authorized licensed use limited to: CLARKSON UNIVERSITY LIBRARY. Downloaded on February 01,2022 at 16:31:54 UTC from IEEE Xplore. Restrictions apply.

Fig. 2. Sample frames collected from databases in the literature and processed.

P24-Rec1-2009), MMI (M&M Initiative) [36], (videos: S030_007_Hap, S032_008_Hap, S033_006_Hap, S034_004_Hap, S047_007_Hap), and CK+ (The Extended Cohn-Kanade Dataset) [37], (used videos: S022_003, S054_001, S077_002, S104_001, S118_002). To maintain symmetries between genuine and fake smiles, we collected five smile videos from each database. Genuine smiles were collected from NEMO and MAHNOB databases where participants' smiles were elicited by showing a sequence of funny or otherwise pleasant video clips. Fake smiles videos were collected from MMI and CK+ databases where participants were asked to perform and instructed by an expert to display a smile.

The collected smile samples were not in the same format, colour, or duration, so we made them as similar as possible. This was done because the samples of CK+ were available as a sequence of frames, whereas the others were videos but of different lengths. The MATLAB R2015a platform was employed to convert them to mp4 format, grey scale, and duration of 5 seconds each. A sample frame of each smile video, for the main study, is shown in Fig. 2. First, frames from each video were extracted (if required). Second, every frame of each sample was resized, maintaining aspect ratios, and then cropped to uniform size. Third, the face portions were masked out so that observers could only watch the smiling faces and not any backgrounds, and other portions were made black in a frame with aspect ratio 4:3 ($\mathrm{Height} = 336$ pixel, $\mathrm{Width} = 448$ pixel) and resolution of 72 dpi. Finally, the processed frames of each sample were converted back into video samples while normalising them for luminance (128 ALU (Arbitrary Linear Unit)) and contrast (32 ALU) [38]. The videos were used as stimuli to record physiological signals and self-report from each observer. Due to smile time duration being generally in the range 0.5s to 4s [10], we chose the stimulus length of videos to be 5s long, adjusted if required by controlling the frame rates of the videos.

We chose 4 databases to avoid database specific results, and because only two of the databases contain both fake and genuine smiles. We then masked and otherwise controlled for illumination and other properties. The results can be seen in Fig. 2, visually it is clear that they are all very similar in appearance.

2.2 Observers

Twenty-six (11 female, 15 male) healthy, right-handed participants were observers in the main study, with the mean age of $30.7 \pm 6.0 (\mathrm{mean} \pm \mathrm{SD})$ years. All the observers provided written consent, and confirmed that their participation was voluntary. An approval from the Australian National University's Human Research Ethics Committee was granted for the study.

2.3 Experimental Instructions

In order to distinguish genuine from fake smiles, a structured experiment was implemented to record and evaluate the observers' judgements and physiological signals. In this paper, we analyse four physiological signals of each observer during smile stimuli observation in the main study. Two of them (ECG and GSR) were recorded using Neulog sensors (https://neulog.com/) at a sampling rate of 10 Hz, according to the device specifications. ECG signals were recorded with sensors on the wrists of both hands, and left elbow. GSR signals were recorded from the index and middle finger of the left hand. BVP was recorded using an Empatica E4 (https://www.empatica.com/) on the wrist of the left arm at a sampling rate of 64 Hz. The sensors were attached in the specified location according to the manufacturer's suggestions. Eye activities were recorded using The Eye Tribe (https://theeyetribe.com/) remote eye-tracker system with a sampling rate of 60 Hz.

Upon arrival at the laboratory, each observer was instructed to sign the consent form and seated on a static chair, facing a 15.6 inch ASUS laptop in a sound-attenuated, dimly lit, closed room. Sensors were attached to measure their ECG, GSR, and BVP signals. Their chairs were moved forward or backwards to adjust the distance between the chair and eye tracker. A spot was displayed on the monitor, and observers asked to track it at first, for calibrating the eye tracker. Observers were instructed to limit their body movements in order to reduce undesired artefacts in the signals. Observers were given a brief introduction to the experimental procedure before starting the experiment, and after attaching the sensors and calibrating the eye tribe (Table 1). During the experiment, all observers used their right hand for moving the mouse or typing. The stimuli were presented to the observers in a randomised fashion. At the end of each stimulus, the observer was asked to answer the four questions as listed in Table 1 (Q1 - Q4), to indicate whether they judged the smile to be a real or fake smile (a single smile was viewed at a time). The total duration of the experiment was around 10 minutes. After completing the experiment, the sensors were removed, and the observers were thanked for their participation.

The data recorded from two female observers were not analysed due to technical problems, leading to poor signal quality, and un-finished data collection. Hence, the analysis results are based on the responses recorded from 24 observers during the main study. In all cases, the answer to question 3 was 'No' and we computed each observer's verbal response rate from the answer of question 1.

2.4 Signal Processing

Four data sets were made from observers' recorded physiological signals, namely pupillary response (PR), BVP, GSR,

Authorized licensed use limited to: CLARKSON UNIVERSITY LIBRARY. Downloaded on February 01,2022 at 16:31:54 UTC from IEEE Xplore. Restrictions apply.

TABLE 1
Experimental Introduction and Instructions.

Sometimes people show facial expressions of emotions they really feel, and smotimes they display expressions that are faked or posed. A real smile shows genuine or spontaneous emotion, like somebody smiles when they get a present or see something funny. A fake smile includes posed or acted expression, like when somebody smiles because they are asked to.

Your task is to decide whether the faces are showing really felt happy smiles or faked/posed smiles. There are 20 videos, each lasting 5 seconds. We want to know how real or fake you think the smiles are.

You will rate each face smile using the following questions:

Q1. How did this smile look to you?
 a) Real (Spontaneous/Genuine)
 b) Fake (Posed/Acted)
Q2. How do you rate your confidence level?
 a) 60% or Less, **b**) 70%, **c**) 80%, **d**) 90%, or **e**) 100%
Q3. Do you recognise the person in the video?
 a) Yes **b**) No
Q4. If yes (recognised), did this affect your decision?
 a) Yes **b**)

No *Each video will finish playing before response options are shown. You will not be able to go back to replay any video clip or go back to change your responses. So, please watch the video clips carefully and then give your response. If you are ready, please* ***click*** *the mouse to proceed.*

and ECG. To create these data sets, observers' physiological signals related to smile videos were separated for each observer excluding physiological responses incidentally recorded while the observers were answering the questions in Table 1. Due to human nature, the pre-processed physiological signals vary differently for different observers, such as pupillary response varied from around 15 mm to 22 mm for O1 (Observer 1) whereas it is around 14 mm to 26 mm for O2. To reduce between observer differences, the pre-processed signals are normalized to [0, 1] by dividing by the particular observer's maximum value for that observer's particular physiological signal [49]. So the values for each video reflect the degree of reaction, and the minimum and maximum are likely to come from different videos viewed by the same observer. Before normalising, data were moved to the positive axis by adding a constant value to make all values positive.

Generally, physiological signals are affected by small signal fluctuations due to the nature of human bodies, physical movements and so on. To remove undesired noise from normalised physiological signals, a number of processing steps were applied. Baseline drift was removed from normalized ECG signals and 0.5 Hz high pass filter (5 point moving window average) was applied for smoothing and filtering purposes [39]. A 20 point median filter was applied in the case of GSR and BVP signals [33], [40]. In the case of eye pupil data, samples where the pupil was obscured due to blinking were measured as zero by the eye tracking system, and cubic spline interpolation was applied to reconstruct the pupil size [41]. Then, a 10-point Hann moving window average was used to smooth the pupil signal [42]. According to Pamplona et al. [43], the pupillary response varies due to effects caused by lighting, and the pupillary light reflex magnitude changes between different people. Principal component analysis (PCA) had been shown to be effective in separating the effect of changes in luminance from stimulus relevance [28]. This was applied here by subtracting the first principal component from the normalized and smoothed pupil data [25]. This procedure was applied on both eyes' data separately, and then averaged to find a single pupillary response signal for a specific observer. An example of processed physiological signals while watching all smile stimuli is shown in Fig. 3. It is clear particularly in the GSR signals that the observer's physiological signals had generally not returned to some baseline value between videos. This adds some noise to our data, but is more like a real scenario is which people do not get a 'time-out' between emotions shown on other people's faces.

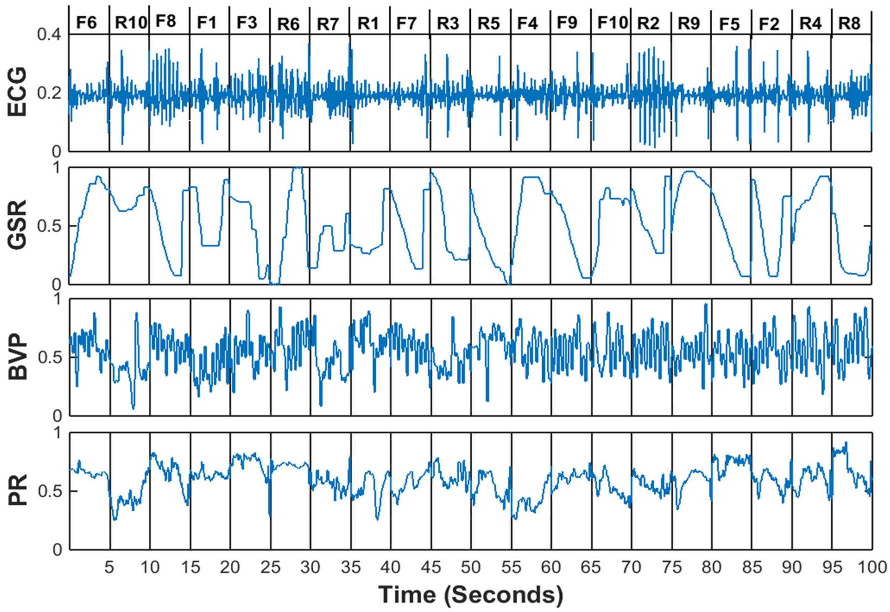

Fig. 3. Processed physiological signals of a single observer during all smile stimuli observations: F/R = Fake/Real.

Authorized licensed use limited to: CLARKSON UNIVERSITY LIBRARY. Downloaded on February 01,2022 at 16:31:54 UTC from IEEE Xplore. Restrictions apply.

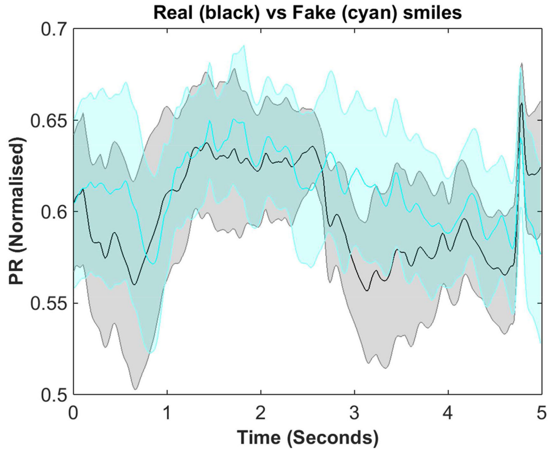

Fig. 4. Average pupillary response timelines over smile stimuli for an observer.

2.5 Feature Extraction

Six temporal features, including amplitude (mean, maximum, minimum), standard deviation, and means of the absolute values of the first and second derivatives of the processed signals, were extracted from each stimulus related physiological signal. These features are very easy to compute and also cover the typical range, gradient and variation of the signals [44]. We computed these features from the first 4.5 seconds of a stimulus relevant physiological signal, instead of over 5 seconds. This is because a ubiquitous high peak is found for most of the PR signals as shown in Fig. 4; a similar pattern is found in [50]. This seemed to reflect the observers' arousal being high at the end of a smile when they were presumably distinguishing real from fake smiles. Overall, there were 6 extracted features for each stimulus and 60 extracted features for all fake smile stimuli, and 60 extracted features for all real smile stimuli, for a particular observer.

2.6 Feature Selection

Feature selection is an important method that reduces large numbers of features by discarding unreliable, redundant, and noisy features, while still achieving comparable, even better performance [44]. In this case we explored a correlation method, named Canonical Correlation Analysis with Neural Network (NCCA), to find an optimal feature set [45]. As features for fake versus real smiles were extracted from the physiological signals of the same observers, there is likely to be an underlying correlation between features for real smiles and features for fake smiles observed. But we believe that there would also be a number of uncorrelated or less correlated features, because observers' physiological behaviour may be different when they were watching real smile stimuli from when they were watching fake smile stimuli. For example, the pupil data follows a common trend for each stimulus. Fig. 4 illustrates that pupils constricted from stimulus onset until reaching a minimum value, then a sharp dilation started and continued until reaching a maximum value. After that point, there were different trends seen according to whether real or fake smile stimuli were being watched. For this purpose, we employed the NCCA system to find minimally correlated feature sets between real and fake smiles. The following joint learning rules (eqs. (1) to (3)) were considered, where i, j, w, s, f, λ and η, η_0 represent observer index, feature index, weight, input features, output features, Lagrange multipliers, and constant learning rates respectively. Initially $\lambda = 0.15, \eta = 0.01, and\ \eta_0 = 0.5$ were chosen and, then weights and Lagrange multipliers were updated according to eqs. (2) and (3). Eq. (1) is a weighted sum over all features of an observer [45]. The NCCA system and PCA were applied only on features of training observers to avoid biasing the test set; this is necessary for leave-one-observer-out cross-validation.

$$f_i = w_i s_i = \sum_j w_{ij} s_{ij} \tag{1}$$

$$\Delta w_{ij} = \eta s_{ij}(f_{i+1} - \lambda_i f_i) \tag{2}$$

$$\Delta \lambda_i = \eta_0 (1 - f_i^2). \tag{3}$$

3 Results and Discussion

We employed the leave-one-observer-out process using two class classifiers to compute classification performance on distinguishing real from fake smiles. We used k-nearest neighbour (KNN), relevance vector machine (RVM), support vector machine (SVM), neural network (NN), and an ensemble over the decisions of these four classifiers. The performance parameters were: 5 nearest neighbours, Gaussian kernel, linear kernel, scaled conjugate gradient training function with 10 hidden nodes, and mean square performance function for KNN, RVM, SVM, NN, and ensemble classifiers respectively. The classification was performed with an Intel(R) Core i5-5200U with 2.20 GHz, 8.00 GB of RAM, Windows 8.1 Operating System 64 bit computer using MATLAB R2015a. The features of one observer were taken as a test set and the rest of the observers' features were used for pre-processing of the training and test set, and then to train the classifier. To remove any effect of bias on the test set, features were selected from correlation analysis on the training set only. This procedure was repeated for each observer, and average accuracies over all observers were reported. The statistical significance of observer data (between real smile stimuli data and fake smile stimuli data) was tested using two-sample Kolmogorov-Smirnov (K-S) test [46]. In a post-hoc analysis, a two-tailed permutation test was performed to find the time points where physiological signals significantly differ over all observers, according to computed p values [47].

According to the K-S test, individual observers' real smile stimuli data significantly differ from fake smile stimuli data considering PR ($p < 0.05$ for all observers), BVP ($p < 0.1$ for O1 & $p < 0.05$ for others), GSR (not significant for O5, O7, O12, O15 & $p < 0.05$ for others), and ECG (not significant for 10 observers & $p < 0.05$ for others) where 'O' denotes an observer. The permutation test finds significant results ($p \leq 0.05$) for a minimum of 14 observers and various significant time points according to the specific physiological signal considered. In the case of PR, mostly lower p values ($p \leq 0.05$) were found between 0.02s to 0.79s, and 7 more segments finishing with 4.34s to 4.40s. On the other hand, mostly lower p values ($p \leq 0.05$) were found in 2.03 s-

Authorized licensed use limited to: CLARKSON UNIVERSITY LIBRARY. Downloaded on February 01,2022 at 16:31:54 UTC from IEEE Xplore. Restrictions apply.

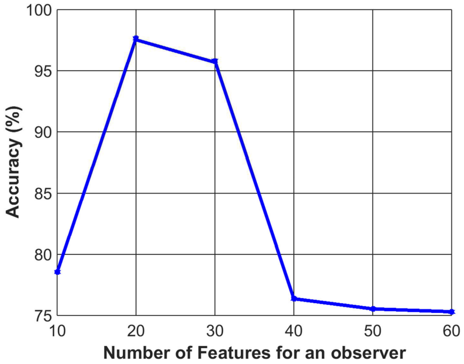

Fig. 5. Variation of accuracies with the number of selected features over all observers.

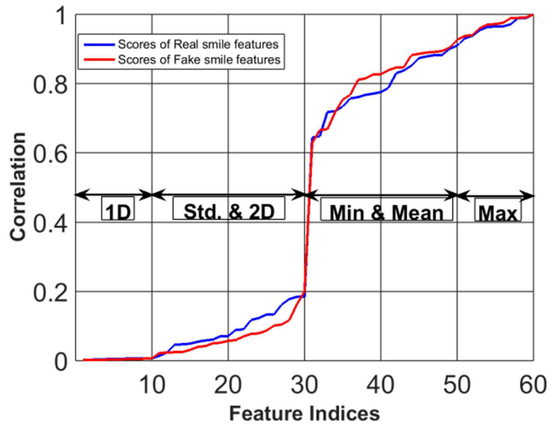

Fig. 6. Features' correlation scores.

2.08 s, and 2.74 s-2.78 s, and 3.42 s-4.12 s for BVP, and GSR, and ECG respectively.

Due to higher numbers of significant results from PR than the other physiological signals, it was used to find the best features sets, as well as the minimum number of observers to reach best accuracies. Fig. 5 shows the classification accuracies of PR with the variation of number of selected features over all observers, reported from the RVM classifier. In total, 1440 PR features (10 stimuli x 6 features x 24 observers) were extracted from all of the fake or real smile stimuli data. Thus, there were overall 60 extracted features for each kind of stimuli for a particular observer. We employed the NCCA system on the extracted feature set to find minimally correlated feature sets according to the RVM classification accuracy. It can be seen from Fig. 5 that higher accuracies (over 95 percent) are found for 20 and 30 selected features.

According to the NCCA system, minimally correlated features were gradients of the signals, followed by variation and range-like characteristics. Mainly, most minimally correlated features were selected from the means of the absolute values of the first differences (1D) followed by standard deviations (Std.) and/or means of the absolute values of the second differences (2D), and then amplitudes (minimum (Min) and/or mean followed by maximum (Max)). The correlation scores of each PR feature are shown in Fig. 6. It can be seen in Fig. 6 that half of the features scored under 0.5: those are mainly gradient and variation like characteristics of observers' physiological signals. Each video has the same number of features selected in Fig. 5; the ordering of features comes from the NCCA selection.

Thus the individual features' meaning is a reflection of similarity to the group of real/fake videos to the training set. In practice this is very similar to retaining a specific column of input values, but NCCA reorders columns so the most important feature is in column one and so on.

We employed the NCCA system to select minimally correlated features in distinguishing between genuine and fake smiles from observers' four physiological signals using four classifiers separately. Finally, an ensemble decision was made from the voting (using the mean value over all classifiers) of individual classifiers. In order to avoid any biasing effect on the test set during classification, the NCCA system was applied only on the training set. This process was repeated for each observer and thus, a leave-one-observer-out process was completed to compute the classification accuracies. In this process, classifiers' training sets were formed from 23 training observers' selected features for a specific physiological signal. The test set features were chosen according to the training observers' features; such as in the case of 30 selected features, the NCCA system selects 1D, Std., and 2D (Fig. 6) for the training set and the test set was formed from these features and also for that test observer. In this way, one observer's features were presented as a test observer while others were used to training the classifiers, and the classifiers decide how accurate the test observer is, to detect real and posed smiles. Then we move to another observer, and the classifier makes a decision. It is important to note that we presented each observer separately (not each video separately, i.e., 10 real videos together and 10 fake videos together (Fig. 7)) and the classifier computed the correctness of that observer. It is also important to note that the selected features do not depend on individual videos, but on the computed temporal features as shown in Fig. 6 (such as in case of 30 selected features, the NCCA system selects three specific features from each video and these are 1D, Std., and 2D). In this way the physiological signals recorded for 24 such observers were presented to the classifier and 24 such decisions were made, and the final accuracy is averaged over all 24 observers. The average classification accuracies over observers are shown in Table 2. It can be seen from Table 2 that higher accuracies with lower standard deviations are found from the ensemble classifier compared to the individual classifiers. It can also be seen that the best result is from the ensemble results

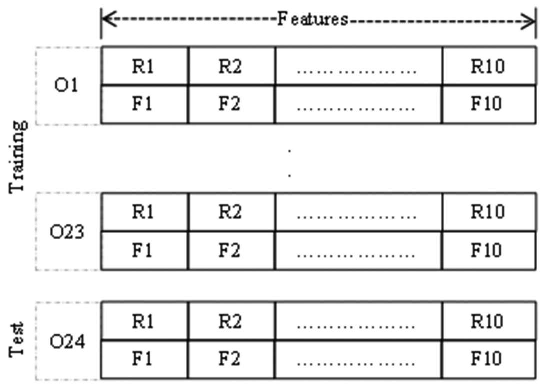

Fig. 7 The leave-one-observer-out classification process; *O* denotes observers, *R* denotes real smile video, and *F* denotes fake smile videos.

Authorized licensed use limited to: CLARKSON UNIVERSITY LIBRARY. Downloaded on February 01,2022 at 16:31:54 UTC from IEEE Xplore. Restrictions apply.

TABLE 2
Average (Avg.) Accuracies and Standard Deviations ($\pm$Std.) Over All Observers.

Signals	No. of Features	KNN		SVM		RVM		NN		Ensemble	
		Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
ECG	20	79.9	13.6	86.1	09.8	78.2	02.3	69.9	15.0	88.5	05.5
	30	79.7	12.1	91.1	08.4	76.3	01.7	91.0	13.8	**91.7**	**03.9**
	40	78.8	11.0	91.0	08.1	75.5	00.9	85.1	13.2	90.8	04.1
BVP	20	76.6	14.8	85.0	11.2	84.6	10.0	81.7	12.5	90.2	05.9
	30	83.2	13.2	93.6	05.6	86.8	07.0	91.9	06.1	**94.5**	**03.5**
	40	82.7	12.3	92.7	02.5	81.6	04.0	90.0	11.7	93.3	03.7
GSR	20	90.7	06.7	97.1	05.0	75.8	01.3	92.4	10.1	95.7	02.3
	30	95.0	05.1	97.5	00.9	75.0	0.01	94.6	10.4	**97.0**	**01.5**
	40	94.5	06.5	97.3	01.1	75.0	0.01	93.5	10.9	96.8	01.7
PR	20	92.7	03.3	97.0	01.1	97.6	01.8	97.2	01.3	97.7	01.0
	30	97.7	02.1	98.0	01.0	95.8	02.3	98.3	01.3	**98.8**	**00.8**
	40	96.1	03.5	97.9	01.1	76.4	00.3	97.5	01.9	97.2	00.9

for observers' PR with 98.8 percent (± 0.8) accuracy. Clearly, the pupil is a good predictor to distinguish displayers' real from fake smiles, and means that "*the eye tells the truth*", or in this case shows true recognition.

We also checked the effect of varying the number of observers used in training on classification accuracies at distinguishing real from fake smiles. As the number of training observers increases, classification accuracy increases for lower number of observers as shown in Fig. 8. After a certain point, the accuracy does not increase much; here '9' is that point. This outcome is reported from PR features and the RVM classifier with the leave-one-observer-out process. For each test observer, 'n', other randomly chosen observers were used to train the classifier and this process was repeated 5 times, with the average outcome shown in Fig. 8.

In order to be able to compare classifier performance with human judgements, we asked each observer for their verbal responses, where the observer gives an opinion on genuine / fake for each specific video stimulus. The percentage of stimuli correctly selected by each observer was computed, and the average correct verbal response over all observers was found to be 59.0 percent in the main study. This verbal accuracy improved to 70.0 percent by voting (50 percent or more of the observers select a smile classification correctly; i.e., more than 12 observers selected 14 correctly $(14/20 * 100 = 70.0\%)$. It can be seen from Fig. 9 that observers' physiological based temporal features are more accurate (more similar to other observers' physiological signals) than their own verbal responses (self-reports).

In the literature, Frank et al. [10] and Hoque et al. [48] found that observers were 56.0 and 69.0 percent correct respectively at discriminating genuine from fake smiles in their experiments. In affective computing so far, the majority of research in this area is done by analysing displayers' smiles directly. As an example, classification accuracies were reported as 91.7 [16], 92.9 [1], 93.0 [15], and 94.0 percent [13], according to their own computational techniques, applied to displayers' image/video based smiles. We also requested Dr. Hamdi Dibeklioğlu to use their technique [1] on our selected videos, without providing our labels to him. When we analysed his classifications, we found that his technique provided 95 percent correctness when applied to our smile videos. The comparative results are shown in Table 3.

It makes intuitive sense from these comparisons and Table 3 that the computational approaches try to use all the information available in the image/video, which is the same information available to the human observer. A slightly better performance by the human physiological signals could be due to greater amount of training available to the human observer from their life prior to the experiment.

Fig. 8. Variation of accuracy with increasing number of training observers.

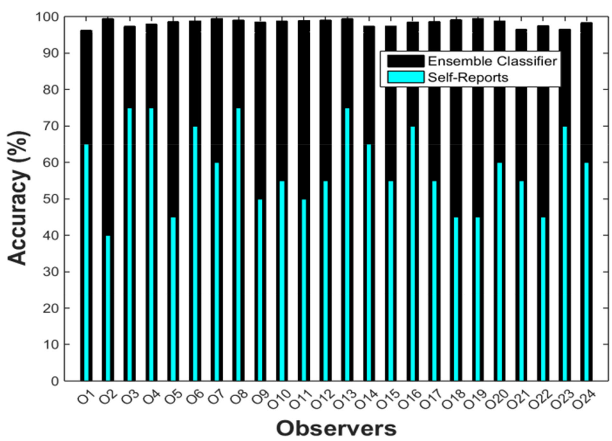

Fig. 9. Comparison between physiological based temporal features and self-reports of each observer at distinguishing real from fake smiles, here 'O' denotes each observer.

Authorized licensed use limited to: CLARKSON UNIVERSITY LIBRARY. Downloaded on February 01,2022 at 16:31:54 UTC from IEEE Xplore. Restrictions apply.

TABLE 3
Comparative Analysis (VR = Verbal Response; CV = Computer Vision

VR	CV [1]	ECG	BVP	GSR	PR
70.0	95.0	91.7	94.5	97.0	98.8

Thus, observing the time course of the human physiological signals, and correlating them to the features of the video that participant was looking at using the eye gaze location data we have collected but not used in this study, could be used to improve computer vision approaches [1], [13], [15], [16] to equal human performance.

In an additional study, to demonstrate that using multiple databases did not introduce some spurious bias, we recorded and analysed BVP and PR signals from 16 new observers $(\mathrm{age} = 21.4 \pm 2.8)$ by showing 5 more fake and 5 more real smile videos (Fig. 10), collected from the UvA-NEMO database [34], (we used videos: 531_deliberate_smile_2, 542_deliberate_smile_2, 559_deliberate_smile_2, 562_deliberate_smile_1, 564_deliberate_smile_2, 002_sponteneous_smile_1, 003_sponteneous_smile_1, 018_sponteneous_smile_1, 020_sponteneous_smile_2, 069_sponteneous_smile_2). We *did not use* any pre-processing steps on smile videos in this case as the videos in a single dataset are the same, but following similar procedures to collect verbal responses and select important features from BVP and PR as describe in Section 2. In this case, observers were found to be 64.4 percent (on average) to 90 percent (by voting) correct by their verbal responses. It is worth mentioning that this experimental paradigm is different from the previous one (observers gained course credit for this study where observers participated voluntarily for previous one), but their average verbal responses of 64.4 percent are quite similar as also reported in [10] and [48]. Using similar procedures of the NCCA system, we select 10, 15, and 20 features from each observer's real/fake smiles' BVP/PR signals as there were in total 30 extracted features for all real or all fake smile stimuli, for a particular observer. The average classification accuracies using a leave-one-observer-out process (as before) over these new 16 observers were computed and depicted in Table 4. It can be seen from Table 4 that the best result (higher accuracy with lower standard deviation) is found from the ensemble results for observers' PR with 98.5 percent (± 01.1) accuracy, essentially the same result as found in our main study, thus indicating that our results are not due to some spurious specific property of the videos or an artefact from our pre-processing but due to the nature of the stimuli. This is also corroborated by some initial work on anger, showing again a similar pattern of results [51]. The results of the experiment show that we can classify real and fake smiles via observers' innate and non-conscious physiological responses that are controlled by the automatic nervous system. There is also evidence that observers' physiological responses can form or evaluate another's mental state [52], perhaps via an internal mimicry which allows us to judge others' facial expression. An important benefit of physiological measurement is that it is not easy to control voluntarily, and provides spontaneous and non-conscious outcomes.

4 Conclusion

In this work, observers' physiological signals were used to distinguish smile displayers' genuine smiles from fake smiles. A number of temporal features were extracted from processed physiological signals. Minimally correlated features, between real and fake smiles, were selected using the NCCA (canonical correlation analysis with neural network) system, from training sets only. All training excluded the physiological signals from the test observer's signals (leave-one-observer-out). Finally, it was seen that observers' PR (pupillary response) features were very good at distinguishing displayer's genuine from fake smiles with a classification rate of 98.8 percent (in the main study) and 98.5 percent (in the additional study) by an ensemble classifier of our four classifiers.

However, several important challenges remain. One of the strengths of the presented work is that we designed the experiment to identify genuine smiles from fake smiles via observers' physiological signals without manipulating the stimuli material. This suggests a general method to evaluate emotions conveyed from stimuli material, especially those with historical data. It could also allow doctors to better estimate patients' mental state, if their reactions differed from

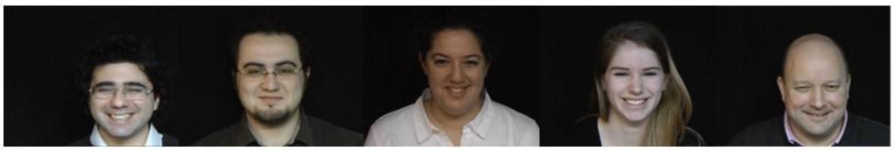

Fig. 10. Sample frames from NEMO database used for additional analysis.

TABLE 4
Average (Avg) Accuracies and Standard Deviations (±Std) Over New 16 Observers on New Sets of Smiles

Signals	No. of Features	KNN		SVM		RVM		NN		Ensemble	
		Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
BVP	10	81.6	11.6	82.6	09.5	87.0	06.3	90.2	09.5	91.6	04.2
	15	86.5	11.7	92.4	08.4	92.3	04.4	93.1	03.0	**94.2**	**03.5**
	20	81.6	14.7	86.3	09.0	87.8	07.7	86.7	10.5	90.1	05.1
PR	10	93.2	04.5	97.5	02.3	96.1	02.2	96.5	02.5	97.6	01.5
	15	93.6	03.5	97.7	01.8	97.1	01.7	98.0	02.5	**98.5**	**01.1**
	20	91.0	09.7	95.6	02.3	76.8	00.4	95.9	11.5	96.4	02.3

Authorized licensed use limited to: CLARKSON UNIVERSITY LIBRARY. Downloaded on February 01,2022 at 16:31:54 UTC from IEEE Xplore. Restrictions apply.

the norm in some clinically relevant manner. In our studies, observers were required to sit still in front of a computer screen with limited body movements requested. While this allowed us to study the feasibility of the proposed method for genuine smiles detection with low noise introduced to the physiological measurements, such requirements on body movements are less likely to be possible in realistic environments. Therefore, future work should extend to more realistic environments, where observers perform daily movements such as gesturing and talking, and even walking.

In addition, the presented method relies on observers' physiological signals, which may vary due to the observers' states. For example, there is evidence that observers' own different facial expressions can affect their visual input processing and that their responses to neutral and happy faces are significantly different [53], in our example this would mean a smiling observer may have different perceptions than a non-smiling observer. While it remains unclear whether or not observers' facial expressions will affect their physiological signals to different types of smiles, future work should be extended to examine how observers' own different facial expressions affect the accuracy of the presented work. This will also make the presented work less vulnerable to observers' individual differences, making it more suitable for real life usage.

The smile stimuli also have some limitations. In the presented work, we masked the background for each smile video, with only the faces of smilers shown. While this allowed us to study the feasibility of the proposed method with low noise caused by side effects of background cues or light intensity effects, it is time consuming and requires a certain amount of manipulation of the original stimuli material. Therefore, future work can investigate the validity of our methods for stimuli which are not masked and are more general, in realistic settings. Further, the duration of our smile videos is short, only lasting for five seconds. This period is just the duration of the smile itself with no pre/post context, being the equivalent of the visual masking in a temporal context. It is also worth examining if the proposed work can be applied or be improved to recognise genuine smiles in longer situations. In addition, we adjusted the backgrounds for the main study but did not adjust for the additional study. There is also evidence that without controlling the background and other parameters, it is possible to differentiate between acted anger and real anger from observers' physiological signals at higher classification accuracy [59]. In the future, we will extend our work with a broader team, such as psychologists, empirical scientists, ethnologists, and so on, with an aim to achieve real world validation of our results, outside controlled laboratory environments. When we will test our system in more realist environments, our participants would see the gesturing and talking of the observers, whose interpretation may also be impacted by cultural differences.

A social good use of our system which we envisage in the future is that our system may be applied to design sensing technologies from care-givers' physiological signals, to record their emotional reactions while caring for disabled people and older people. For example, consider the usefulness of the Assistive Context Aware Toolkit (ACAT), which was designed for Stephen Hawking to enable him to control his computer and communicate with others [60] – our system could help his ACAT system to interact with care givers, by recognising their emotional reactions, essentially modelling (and perhaps even responding to) the unconscious emotional interplay between people in conversations. Our system can also be applied to measure past emotions in historical data, such as measuring the genuineness of the facial expression of Adolf Hitler by using observers watching his videos, or answering security questions, such as "does Kim Jong-un believe what he is saying" or "do people believe him"? These speculations can only be tested after our work has been extended and verified beyond laboratory settings.

We should also consider the limitations on our work in respect of the nature of the data we have used. The physiological signals and the time-scales at which we have used them clearly depend on fast connections to the autonomic nervous system, and these 'instinctive' responses are not guaranteed to be correct in real social situations. This means that the next step is not to produce a wrist-borne device which will tell us "you did feel that was a real smile" as this may not work at all in real social situations, and may raise ethical and legal issues. For example, is it a breach of privacy if a program automatically recognises a user's instinctive reaction?

Overall, we have produced excellent results by using sensors on the viewer of a smile rather than the producer of a smile. This means our results can be extended to historical data as mentioned before, and to non-human data, for example to examine the veracity of smiles of virtual avatars [54], [55] as it is clear that the brain areas controlling expression recognition and creation interact without conscious control [56] hence knowing it is an avatar may not change physiological signals caused by the recognition / creation pathways. We intend to investigate this in our future work, by comparing avatars constructed from genuine smile videos and from posed smile videos. We will also consider dynamic / recurrent versions of neural networks such as the Elman recurrent neural network in the future, to potentially improve the recognition in longer duration settings which may be more robust to sensing artefacts related to the physiological signals and measurement devices.

We believe that the limited social contexts of computerised avatar use so far and in the near future avoid the limitations of our work identified above. Thus our work could be used to improve avatars, such as Nadia [57], intended to help users access the National Disability Insurance Scheme in Australia [58]. Nadia will be less suitable for purpose if its users' instinctive reaction is "this is not a real smile".

Acknowledgments

The authors wish to thank all observers for their participation and taking part in this research. We are grateful to the anonymous reviewers for their help especially in contextualising our work – thank you.

References

[1] H. Dibeklioğlu, A. A. Salah, and T. Gevers, "Recognition of genuine smiles," *IEEE Trans. Multimedia*, vol. 17, no. 3, pp. 279–294, Mar. 2015

Authorized licensed use limited to: CLARKSON UNIVERSITY LIBRARY. Downloaded on February 01,2022 at 16:31:54 UTC from IEEE Xplore. Restrictions apply.

[2] C. Darwin, *The Expression of the Emotions in Man and Animals*. New York, NY, USA: Philosophical Library, 1998 (Original work published, 1872).

[3] R. S. Woodworth, *Experimental Psychology*, New York, NY, USA: Henry Holt, 1938.

[4] S. S. Tomkins and R. M. C. Carter, "What and where are the primary affects? Some evidence for a theory," *Perceptrial Motor Skills*, vol. 18, pp. 119–158, 1964.

[5] D. Keltner, T. E. Moffitt, and M. Stouthamer-Loeber, "Facial expressions of emotion and psychopathology in adolescent boys," *J Abnorm Psychol.*, vol. 104, no. 4, pp. 644–652, 1995.

[6] D. Keltner and P. Ekman, "Facial expression of emotion," *Handbook of Emotions*, New York, NY, USA: Guilford Publications, Inc., 2000, pp. 236–249.

[7] N. Eisenberg, R. A. Fabes, P. A. Miller, J. Fultz, R. Shell, R. M. Mathy, and R. R. Reno, "Relation of sympathy and personal distress to prosocial behavior: A multimethod study," *J. Personality Soc. Psychol.*, vol. 57, no. 1, pp. 55–66, 1989.

[8] F. Esteves, U. Dimberg, and A. Öhman, "Automatically elicited fear: Conditioned skin conductance responses to masked facial expressions," *Cognition Emotion*, vol. 8, no. 5, pp. 393–413, 1994.

[9] D. Keltner, R. C. Young, B. N. Buswell, "Appeasement in human emotion, social practice, and personality," *Aggressive Behavior*, vol. 23, pp. 359–375, 1997.

[10] M. G. Frank, P. Ekman, and W. V. Friesen, "Behavioral markers and recognizability of the smile of enjoyment," *J. Personality Soc. Psychol.* vol. 64, no. 1, pp. 83–93, 1993.

[11] P. Ekman, R. J. Davidson, and W. V. Friesen, "The Duchenne smile: Emotional expression and brain physiology: II," *J. Personality Soc. Psychol.*, vol. 58, pp. 342–353, 1990.

[12] P. Ekman, W. V. Friesen, and M. O'Sullivan, "Smiles when lying," *J. Personality Soc. Psychol.*, vol 54, no. 3, pp. 414–420, 1988.

[13] M. F. Valstar, H. Gunes, and M. Pantic, "How to distinguish posed from spontaneous smiles using geometric features," in *Proc. ACM Int. Conf. Multimodal Interfaces*, Nov. 2007, pp. 38–45.

[14] Z. Ambadar, J. F. Cohn, and L. I. Reed, "All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous," *J. Nonverbal Behav.*, vol. 33, no. 1, pp. 17–34, 2009.

[15] J. F. Cohn and K. L. Schmidt, "The timing of facial motion in posed and spontaneous smiles," *Int. J. Wavelets Multi-Resolution Inf. Process.*, vol. 2, pp. 1–12, 2004.

[16] Q. Gan, C. Wu, S. Wang, and Q. Ji, "Posed and spontaneous facial expression differentiation using deep Boltzmann machines," in *Proc. Int. Conf. Affective Comput. Intell. Interaction*. 2015, pp. 643--648.

[17] U. Hess and R. E. Kleck, "The cues decoders use in attempting to differentiate emotion-elicited and posed facial expressions," *Eur. J. Soc. Psychol.*, vol. 24, no. 3, pp. 367–381, 1994.

[18] N. A. Murphy, J. M. Lehrfeld, and D. M. Isaacowitz, "Recognition of posed and spontaneous dynamic smiles in younger and older adults," *Psychol. Aging.*, vol. 25, no. 4, pp. 811–821, 2010.

[19] P. Ekman and W. V. Friesen, "Felt, false, and miserable smiles," *J. Nonverbal Behav.*, vol. 6, pp. 238–252, 1982.

[20] R. W. Picard, *Affective Computing*, Cambridge, Massachusetts, London, England: MIT Press, 2000, pp. 175.

[21] J. Kim and E. Andre, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2067–2083, Dec. 2008.

[22] M. Soleymani, G. Chanel, J. J. M. Kierkels, and T. Pun, "Affective characterization of movie scenes based on content analysis and physiological changes," *Int. J. Semantic Comput.*, vol. 3, no. 2, pp. 235–254, Jun. 2009.

[23] J. Kim, "Bimodal emotion recognition using speech and physiological changes," in *Robust Speech Recognition and Understanding*, M. Grimm and K. Kroschel, Eds., Vienna, Austria: I-Tech Education and Publishing, 2007, pp. 265–280.

[24] A. R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY, USA: Grosset/Putnam, 1994.

[25] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Trans. Affective Comput.*, vol. 3, no. 2, pp. 211–223, Apr.-Jun. 2012.

[26] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, 'The pupil as a measure of emotional arousal and autonomic activation," *Psychophysiology*, vol. 45, no. 4, pp. 602–607, 2008.

[27] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *Int. J. Hum.-Comput. Stud.*, vol. 59, pp. 185–198, 2003.

[28] F. T. P. Oliveira, A. Aula, and D. M. Russell, "Discriminating the relevance of web search results with measures of pupil size," *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2009, pp. 2209–2212.

[29] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, 2010.

[30] R. W. Levenson, P. Ekman, and W. V. Friesen, "Voluntary facial action generates emotions-specific automatic nervous system activity," *Psychophysiology*, vol. 27, pp. 363–384, 1990.

[31] M. M. Bradley, M. Codispoti, B. N. Cuthbert, and P. J. Lang, "Emotion and motivation I: Defensive and appetitive reactions in picture processing," *Emotion*, vol. 1, no. 3, pp. 276–298, 2001.

[32] J. Healey and R. Picard, "Digital processing of affective signals," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 1998, pp. 3749–3752.

[33] A. Kushki, J. Fairley, S. Merja, G. King, and T. Chau, "Comparison of blood volume pulse and skin conductance responses to mental and affective stimuli at different anatomical sites," *Physiological Measures*, vol. 32, no. 10, pp. 1529–1539, 2011.

[34] H. Dibeklioglu, A. A. Salah, and T. Gevers, "Are you really smiling at me? Spontaneous versus posed enjoyment smiles," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 525–538.

[35] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affective Comput.*, pp. 42–55, vol. 3, no. 1, Jan.-Mar. 2012.

[36] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," *Proc. IEEE Int. Conf. Multimedia Expo*, 2005, Art. no. 5.

[37] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete expression dataset for action unit and emotion-specified expression," *Proc. 3rd Int. Workshop CVPR Hum. Commun. Behav. Anal.*, 2010, pp. 94–101.

[38] V. Willenbockel, J. Sadr, D. Fiset, G. O. Horne, F. Gosselin, and J. W. Tanaka, "Controlling low-level image properties: the SHINE toolbox," *Behav. Res. Methods*, pp. 671–684, vol. 42, no. 3, 2010.

[39] O. AlZoubi, S. K. D'Mello, and R. A. Calvo, "Detecting naturalistic expressions of nonbasic affect using physiological signals" *IEEE Trans. Affective Comput.*, pp. 298–310, vol. 3, no. 3, Jul.-Sep. 2012.

[40] R. Guo, S. Li, L. He, and W. Gao, "Pervasive and unobtrusive emotion sensing for human mental health," in *Proc. 7th Int. Conf. IEEE Pervasive Comput. Technol. Healthcare*, May 5-8, 2013, pp. 436–439.

[41] S. Mathôt, "A simple way to reconstruct pupil size during eye blinks," 2013. [Online]. Available: http://dx.doi.org/10.6084/m9.figshare.688001.

[42] W. L. Zheng, B. N. Dong, and B. L. Lu, "Multimodal emotion recognition using EEG and eye tracking data," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 5040–5043.

[43] V. F. Pamplona, M. M. Oliveira, and G. V. G. Baranoski, "Photorealistic models for pupil light reflex and iridal pattern deformation," *ACM Trans. Graph.*, vol. 28, no. 4, pp. 1–12, 2009.

[44] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001.

[45] M. Z. Hossain, M. M. Kabir, and M. Shahjahan, "A robust feature selection system with Colin's CCA network," *Neurocomputing*, vol. 173, no. 3, pp. 855–863, 2016.

[46] G. Marsaglia, W. Tsang, and J. Wang, "Evaluating Kolmogorov's distribution," *J. Statist. Softw.*, vol. 8, no. 18, 2003.

[47] R. C. Blair and W. W. Karniski, "An alternative method for significance testing of waveform difference potentials," *Psychophysiology*. Hoboken, NJ, USA: Wilay, 1993, pp. 518–524.

[48] M. E. Hoque, D. J. McDuff, and R. W. Picard, "Exploring temporal patterns in classifying frustrated and delighted smiles," *IEEE Trans. Affective Comput.*, vol. 3, no. 3, pp. 323–334, Jul.-Sep. 2012.

[49] M. Z. Hossain, T. Gedeon, and R. Sankaranarayana, "Observer's galvanic skin response for discriminating real from fake smiles," in *Proc. 27th Australasian Conf. Inf. Syst.*, 2016, pp. 1–8.

[50] M. Z. Hossain, T. Gedeon, R. Sankaranarayana, D. Apthorp, and A. Dawel, "Pupillary responses of Asian observers in discriminating real from fake smiles: A preliminary study," in *Proc. 10th Int. Conf. Methods Techn. Behav. Res.*, 2016, pp. 170–176.

[51] L. Chen, T. Gedeon, M. Z. Hossain, and S. Caldwell, "Are you really angry? Detecting emotion veracity as a proposed tool for interaction," in *Proc. 29th Australian Conf. Hum.-Comput. Interaction*, 2017, Art. no. 5.

Authorized licensed use limited to: CLARKSON UNIVERSITY LIBRARY. Downloaded on February 01,2022 at 16:31:54 UTC from IEEE Xplore. Restrictions apply.

[52] P. Shah, C. Catmur, and G. Bird, "From heart to mind: Linking interoception, emotion, and theory of mind," *Cortex*, vol. 93, pp. 1–4, 2017.
[53] A. Sel, B. Calvo-Merino, S. Tuettenberg, and B. Forster, "When you smile, the world smiles at you: ERP evidence for self-expression effects on face processing," *Soc. Cogn. Affect. Neurosci.*, vol. 10, no. 10, pp. 1316–1322, 2015.
[54] M. Seymour, K. Riemer, and J. Kay, "Interactive realistic digital avatars - Revisiting the uncanny valley," in *Proc. HICSS-50*, 2017, pp. 547–556.
[55] M. Z. Hossain and T. Gedeon, "Discriminating real and posed smiles: Human and avatar smiles," in *Proc. 29th Australian Conf. Human-Comput. Interaction*, 2017, pp. 581–586. https://doi.org/10.1145/3152771.3156179.
[56] K. M. Gothard, "The amygdalo-motor pathways and the control of facial expressions," *Frontiers Neurosci.*, vol. 8, 2014, Art. no. 43.
[57] Soul Machines, Level 1, 106 Customs Street West, Auckland 1010, New Zealand, https://www.soulmachines.com/
[58] DHS, Australian Government Department of Human Services (2018). Digital Delivery of Government Services, Submission to the Finance and Public Administration References Committee, 17 October 2018, https://tinyurl.com/dhs-nadia, Page 16.
[59] L. Chen, T. Gedeon, M. Z. Hossain, and S. Caldwell, "Are you really angry?: detecting emotion veracity as a proposed tool for interaction," in *Proc. 29th Australian Conf. Comput.-Hum. Interaction*, 2017, pp. 412–416.
[60] P. Denman, L. Nachman, and S. Prasad, "Designing for "a" user: Stephen Hawking's UI," in *Proc. 14th Int. Proc. Participatory Des. Conf.*, 2016, pp. 94–95.

Md Zakir Hossain received the BSc and MSc degrees both from the Department of Electrical and Electronic Engineering (EEE), Khulna University of Engineering & Technology (KUET), Bangladesh. He is currently working toward the PhD degree in Human-Centred Computing (Hcc) group of the Research School of Computer Science at The Australian National University, Australia. He is now on leave from the Department of EEE, KUET, Bangladesh where he was serving as an assistant professor. His research interests include affective computing, information retrival, and biomedical signal processing.

Tom Gedeon received the BSc (Hons) and PhD degrees from the University of Western Australia. He is currently chair professor of computer science at the Australian National University, Canberra, Australia, and leads the Human-Centred Computing Group at the Research School of Computer Science. His research interests include bio-inspired computing and in human centred computing. He is a former president of the Asia-Pacific Neural Network Assembly and a former president of the Computing Research and Education Association of Australasia. He serves on journal advisory boards as member or editor. He is a senior member of the IEEE.

Ramesh Sankaranarayana received the ME degree from the Indian Institute of Science, India, and the PhD degree from the University of Alberta, Canada. He is currently the associate director (Educational partnerships), Research School of Computer Science, Australian National University, Canberra, Australia. He research interests include information retrieval, human-centered computing and software engineering. He is a member of the ACM.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.

Authorized licensed use limited to: CLARKSON UNIVERSITY LIBRARY. Downloaded on February 01,2022 at 16:31:54 UTC from IEEE Xplore. Restrictions apply.