Observer's Galvanic Skin Response for Discriminating Real from Fake Smiles

Md Zakir Hossain
Research School of Computer Science
The Australian National University
Canberra, Australia
Email: zakir.hossain@anu.edu.au

Tom Gedeon
Research School of Computer Science
The Australian National University
Canberra, Australia
Email: tom@cs.anu.edu.au

Ramesh Sankaranarayana
Research School of Computer Science
The Australian National University
Canberra, Australia
Email: ramesh@cs.anu.edu.au

Abstract

This paper demonstrates a system to discriminate real from fake smiles with high accuracy by sensing observers' galvanic skin response (GSR). GSR signals are recorded from 10 observers, while they are watching 5 real and 5 posed or acted smile video stimuli. We investigate the effect of various feature selection methods on processed GSR signals (recorded features) and computed features (extracted features) from the processed GSR signals, by measuring classification performance using three different classifiers. A leave-one-observer-out process is implemented to reliably measure classification accuracy. It is found that simple neural network (NN) using random subset feature selection (RSFS) based on extracted features outperforms all other cases, with 96.5% classification accuracy on our two classes of smiles (real vs. fake). The high accuracy highlights the potential of this system for use in the future for discriminating observers' reactions to authentic emotional stimuli in information systems settings such as advertising and tutoring systems.

Keywords GSR, Recorded Features, Extracted Features, Observers, Smiles.

1 Introduction

A smile is a complex and multi-purpose facial display that conveys not only the meaning of happiness (Ekman et al. 1990), but can also be identified as frustration, depression, empathetic, surprise, polite disagreement, pain and even more (Hoque et al. 2011). We refer to these latter acted, posed and non-happy smiles as fake smiles, and refer to smiles related to happiness as real smiles. In this scenario, it is important to accurately discriminate real from fake smiles to understand the affective state underlying the meaning of smiles. In the past, researchers measured displayers' smiling characteristics to discriminate real from fake smiles (Ambadar et al. 2009; Hoque et al. 2011), rather than using observers' reactions. In our work, we analyse an observer's galvanic skin response (GSR) while they watch a (video of a) displayer's smile.

A GSR signal indicates electrical changes measured at the surface of human skin that varies with the changes in skin moisture level (sweating) and reflects the changes in their sympathetic nervous system (Nourbakhsh et al. 2012). It is an automatic reaction which cannot be controlled by the user and is considered one of the strongest signals that can be used for emotion detection (Geo et. al. 2013; Nakasone et al. 2005). GSR signals can also be used for stress or cognitive load detection to improve mental health (Engstrom et al. 2005). To the best of our knowledge, no one has previously investigated *observers'* GSR signals to discriminate real from fake smiles. In this paper, we perform a preliminary study on 10 observers' GSR signals, while they are watching fake and real smiling video stimuli.

Due to the nature of human bodies, the raw GSR signals are affected by noise, such as body movements, small fluctuations etc. (Geo et al. 2013). To remove the noise, raw GSR signals are normalized and filtered; we refer to these as recorded features. On the other hand, feature extraction is an important technique before measuring accuracy (Picard et al. 2001). Thus, seven features are extracted from each processed GSR signal; we refer to these as extracted features. We found that all recorded and extracted features are not equally important for discriminating real from fake smiles. Thus feature selection methods are applied to find important features by removing unreliable and noisy features, while still achieving comparable and even better performance. In this case, five feature selection methods (MI = Mutual Information, SD = Statistical Dependency, RSFS = Random Subset Feature Selection, SFS = Sequential Forward Selection, and SFFS = Sequential Floating Forward Selection) are employed to select important and informative features from recorded and extracted features respectively (Jouni et al. 2015). Finally, three separate classifiers (KNN = K-nearest Neighbour, SVM = Support Vector Machine, and NN = Neural Network) are implemented to compute the classification performance on data from one observer, while the features from other observers are used to train the classifier. Thus, a leave-one-observer-out process is performed and results are reported as average values over all observers. A sketch of the process is in Figure 1.

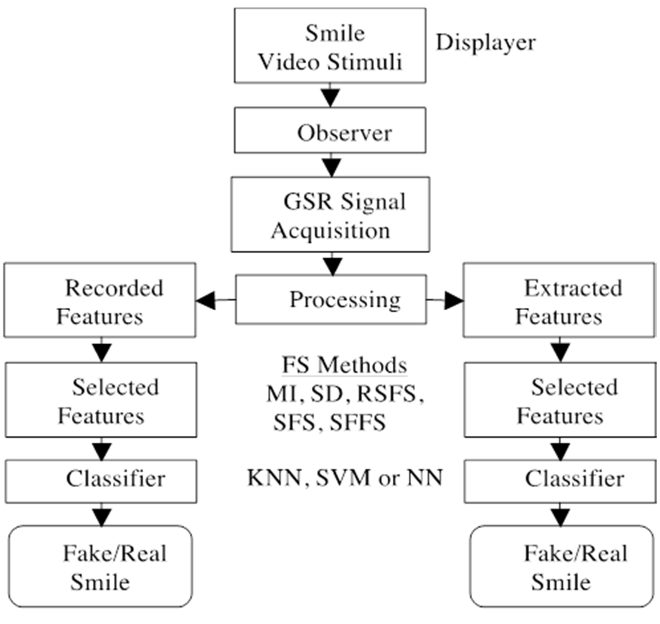


Figure 1: Research methodology. FS = Feature Selection, MI = Mutual Information, SD = Statistical Dependency, RSFS = Random Subset Feature Selection, SFS = Sequential Forward Selection, SFFS = Sequential Floating Forward Selection, KNN = K-nearest Neighbours, SVM = Support Vector Machine, and NN = Neural Network.

Information Systems are influenced by studies on user behaviour intention, psychology, affect detection, and so on (Sun and Zhang 2006; Willis and Jones 2012). This research focuses on determining how an observer's GSR reflects their understanding of a displayer's smile information, to discriminate real from fake smiles. It motivates objective understanding of how individuals' GSR responds to fake vs. real smiles that they observe in typical environments. Recording GSR from observers could be used for purposes beyond our study. For example, clinicians use observable nonverbal and other uncontrollable behaviours to assess patient's affective state (Kasl and Mahl 1965). Police use GSR and other relatively uncontrollable physiological responses for lie detection (Lykkenn 1974). It may also be possible to detect from a Customs' officer's GSR whether the person at passport control appears suspicious or not.

We use observers and record their signals to allow for some general use of our approach in social/emotional settings. We are interested in the perception of observers (Kraut 1978) when observing people (e.g. smiles) or their data on a screen. This has wide potential applicability to Information System research.

2 Experiment

2.1 Aim

The goal of the present study is to find a classifier and a feature selection method to best use an observer's GSR features to predict the hidden mental state of the person in the video (displayer), discriminating between happy smiles and other kind of smiles. This is important for information systems to support emotion communication (Willis and Jones 2012).

2.2 Observers

Ten graduate students, 25 to 39 years old, participated as observers (participants) in this experiment. 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.

2.3 Video Stimuli

Twenty-five smiling video clips are selected from three benchmark databases (10 from AFEW (Dhall et al. 2014), 5 from MMI (Pantic et al. 2005) and 10 from MAHNOB (HCI and Laughter) (Petridisa et al. 2013; Soleymani et al. 2012)) to use as stimuli. The AFEW stimuli are from a collection of acted smiles by professional actors. The MMI stimuli are identified as fake smiles, because displayers were asked and instructed to display a smile (Valstar and Pantic 2010). On the other hand, in the case of MAHNOB, the displayers' smiles were elicited by watching a sequence of funny or pleasant video clip (Soleymani et al. 2012) stimuli and treated here as genuine/real smiles. The collected smiling video stimuli are cropped to keep the face portion only, and processed using MATLAB R2015a to make them identical in size, format and duration. These were: grey scale, mp4 format and lasting 10 seconds each. This paper is based on analysing observers' GSR features, while watching 10 smiling stimuli (5 of MMI and 5 of MAHNOB_HCI); the others are not considered due to technical problems, avoiding famous actors' smiles (perhaps they are famous because they can do realistic fakes?), and lack of explanation of the stimuli in the source.

2.4 Data Acquisition

To collect GSR signals, a Neulog (https://neulog.com/) sensor is attached to the observers' index and middle finger of the left hand. The sampling frequency was 10 Hz. A 17.1" LCD monitor and a usual computer mouse are peripherals for interaction between observer and a PC running the web-based smiling video stimuli. Each stimulus is followed by a five point (-2 to +2) Likert Scale where negative and positive ends of the scale allow selection of the video stimuli as fake and real smiles respectively, with higher or lower confidence. The middle point represents confusion.

2.5 GSR Signal Processing

To reduce the between-observer differences, a normalization technique is applied on the collected raw GSR signals (Nourbakhsh et al. 2012). Each value of a particular observer's GSR signal is divided by the maximum value of that observer's GSR signal. Generally, the GSR signal is affected by small signal fluctuations due to the nature of human bodies, physical movements and so on. To remove the undesired noise from normalized GSR signal, a 20 point median filter is applied (Guo et al. 2013). Although recorded features are in the range of [0, 1] at all times, extracted features deviated from this range when we calculated derived values. For reducing data differences the extracted features are rescaled to [0, 1] (if required) (Guo et al. 2013) by dividing by the extracted maximum value.

3 Features

3.1 Recorded Features

In this case, each time point of processed GSR signals were treated as features. Thus, there were 100 recorded features for a stimulus and 500 recorded features for all fake smiling stimuli and 500 for all real smiling stimuli, for a particular observer. During training, only 9,000 features (10 stimuli x 100 features x 9 observers; i.e., from the other 9 observers) are used, with 1,000 for testing.

3.2 Extracted Features

The extracted GSR features include six temporal features and one spectral feature for each stimulus. The six temporal features include mean and standard deviations of the normalized signals, means of the absolute values of the first and second differences of the normalized signals and the filtered signals. These features can be computed in a cost-effective way and cover the typical range, gradient and variation of the signals (Picard et al. 2001). The periodic excitation is another important characteristic of the GSR signal, which can be represented by spectral features (Guo et al. 2013). We examined different variations of power spectrum using Welch power spectrum density in the MATLAB platform. The processed GSR signal of each stimulus is examined to find the frequency contents in that signal. Each stimulus related GSR signal shows about similar variation as shown in Figure 2. It is clear from Figure 2 that the power feature has non-zero values in frequencies less than 1 Hz, and mostly less than 0.5 Hz. In this situation, important features can be found within a 16-datapoint-length (Nourbakhsh et al. 2012). In this regard, the first 16 data-points are used to compute the spectral mean from each stimulus processed GSR signal. Finally, there are 7 extracted features for a stimulus and 35 extracted features for either all fake smiling stimuli or real smiling stimuli for a particular observer. During training, only 630 features (10 stimuli x 7 features x 9 observers; i.e., from the other 9 observers) are used, with 70 for testing.

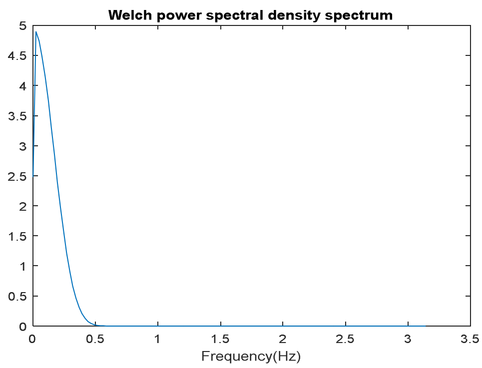


Figure 2: Welch's power spectral density spectrum of first MMI stimulus, Observer 1.

4 Feature Selection

We use MATLAB feature selection code (Jouni et al. 2015) to select important and informative features from recorded and extracted features, respectively. There are five feature selection algorithms as follows.

i. MRMR (Minimal-Redundancy-Maximum-Relevance) - finds features by estimating mutual information (MI) between features and associated class labels with maximizing feature relevance and minimizing redundancy,

ii. SD (Statistical Dependency) - finds features according to measuring the statistical dependency between features and associated class labels,

iii. RSFS (Random Subset Feature Selection) - choses features from the entire feature pool by comparing their relevance values,

iv. SFS (Sequential Forward Selection) - fills empty feature set according to their best scores by removing redundant features, and

v. SFFS (Sequential Floating Forward Selection) - uses backward steps (if required) after each forward step to find better features than previously selected.

In the case of MI and SD, 12 quantization labels are used to select features. Also, the unweighted average recall (UAR) performance criterion and KNN classifier are implemented for relevance update and measuring scores, respectively, in the case of RSFS, SFS and SFFS. We used the feature selection code only on training features (i.e. on nine observers' data) by default KNN classifier as in (Jouni et al. 2015), with 10 fold cross validation to compute the scores of the features. The final feature set is formed by those features that are common to score higher in the maximum number of folded outcome.

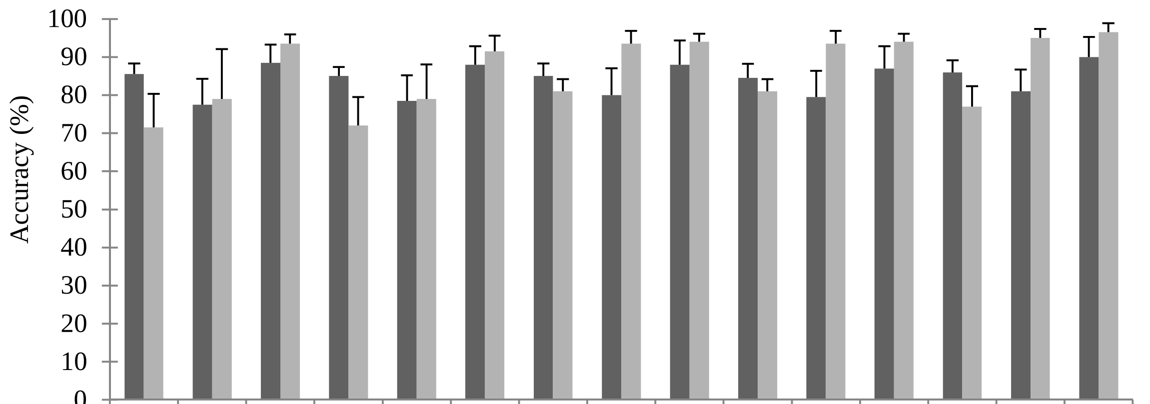
5 Results and Discussion

The performance of five feature selection methods on discriminating real from fake smiles are evaluated with three, two class (fake and real smile) binary classifiers, being k-nearest neighbour (KNN), support vector machine (SVM) and simple neural network (NN). The performance parameters are 3 nearest neighbours, least square method with quadratic kernel function, scaled conjugate gradient training function with mean square error performance function and 5 hidden nodes for KNN, SVM and NN respectively. The program is performed with an Intel(R) Core™ i5-5200U with 2.20 GHz, 8.00 GB of RAM, Operating System 64-bit computer using MATLAB R2015a. A leave-one-observer-out process is implemented to test the classification performance. The features of one observer are taken as a test set and rest of the observers' features are used to train the classifier. To remove the effect of biasing on test set, features are selected from training set only. This process is repeated for each observer. For that reason, the number of selected features is varied according to observer as shown in Table 1. It is notable from Table 1, that a higher percentage of features are selected in the case of extracted features (example - 10 out of 35, 28% selected) than recorded features (example - 36.3 out of 500, 7% selected).

Test Observer	Extracted Features (EF)					Recorded Features (RF)				
	MI	SD	RSFS	SFS	SFFS	MI	SD	RSFS	SFS	SFFS
O1	10	10	8	3	3	50	20	34	4	4
O2	10	10	11	2	2	32	50	18	4	6
O3	10	10	6	3	3	34	34	26	2	4
O4	10	10	11	3	5	50	50	12	4	5
O5	10	10	11	2	3	9	9	62	3	5
O6	10	10	8	3	3	32	32	25	6	8
O7	10	10	9	3	4	38	38	10	4	5
O8	10	10	11	4	3	50	50	30	4	5
O9	10	10	11	5	5	34	34	11	6	7
O10	10	10	9	3	6	34	34	26	4	4
Average	10	10	9.5	3.1	3.7	36.3	35.1	25.4	4.1	5.3

Table 1. Number of selected features for each test observer

The average classification accuracies over observers are reported in Figure 3. The error bars represent standard deviation. We see that, KNN shows higher accuracies in the case of recorded features than extracted features, and the opposite case is true for SVM and NN. We also observed that KNN shows higher accuracies for recorded features than SVM and vice versa for extracted features. This may depend on parameter tuning. We also check that accuracies are increased with decreasing 'k' values of KNN and vice versa. If k=1, then accuracies are increased about 5%. On the other hand, accuracies are decreased about 8-10% for linear kernel of SVM instead of quadratic kernel.



	SFS			SFFS			MI			SD			RSFS		
	KNN	SVM	NN	KNN	SVM	NN	KNN	SVM	NN	KNN	SVM	NN	KNN	SVM	NN
RF	85.5	77.5	88.5	85	78.5	88	85	80	88	84.5	79.5	87	86	81	90
EF	71.5	79	93.5	72	79	91.5	81	93.5	94	81	93.5	94	77	95	96.5

Figure 3: Average classification accuracies over observer, RF=Recorded Features & EF=Extracted Features.

We can observe from Figure 3 that GSR is an effective physiological signal at discriminating real from fake smiles where accuracy is found as high as 96.5%. In this case, NN and RSFS are dominant classifier and feature selection method respectively when compared to others. We also observed that NN takes a long time (2.1 sec to 3.4 sec) to compute accuracy, when compared to SVM (about 0.24 sec) and KNN (0.09 sec to 0.1 sec). On the other hand, RSFS is costly compared to other feature selection methods as shown in Table 2.

	MI	SD	RSFS	SFS	SFFS
RF	0.6127	0.0577	10.8268	9.3705	9.3926
EF	0.0124	0.0122	03.5272	0.5176	0.5565

Table 2. Elapsed time (in sec) during feature selection

It is hard to compare our results with others in the literature. This is because, to the best of our knowledge, no previous research reports observers' GSR signals (or other physiological signal) having been used to discriminate between displayers' real from fake smiles. Palanisamy et al (Palanisamy et al. 2013) used GSR and other physiological signals for stress identification and showed that probabilistic neural networks and KNN were 70.0% and 70.8% correct respectively using only the GSR signal. Guo et al (Guo et al. 2013) used SFFS based GSR features to recognize four emotions using KNN and found 79.5% accuracy. Shangguan et al (Shangguan et al. 2014) employed a curve fitting model and found correct-recognition rate up to 91.4%. Sharma and Gedeon (Sharma and Gedeon 2014) used observers' physiological signals to predict displayer stress with 90% accuracy. These outcomes are not directly comparable with our results due to different scenarios; nevertheless our raw numbers are higher, even though we are arguably solving a harder smile recognition problem, measuring effects on the observer and not the displayer of the smile.

6 Conclusion

In this paper, we experimentally evaluated GSR signals to locate more dominant features for discriminating real from fake smiles from observers of the smiles. It is a pervasive, simple, nonintrusive and easily captured signal. The original GSR signals are smoothed and de-noised; features are extracted and selected, and normalization is applied to reduce the observer-dependency. We investigate both recorded and extracted features to compute classification accuracies using KNN, SVM and NN. Various feature selection methods and their outcomes are investigated. Finally, higher accuracy (96.5%) is found on RSFS based selected features using simple NN compared to others. These findings make it applicable in real world situations, perhaps using mobile applications to monitor the displayers' smiles by observers to improve their emotional intelligence. Our future work will include assessing the other physiological signals by observing different emotional videos, using real world scenarios such as advertising websites (Gregor et al. 2014) to improve classification accuracies, and comparison to verbal response accuracy rate.

7 References

Ambadar, Z., Cohn, J.F., and Reed, L.I. 2009. "All Smiles are Not Created Equal: Morphology and Timing of Smiles Perceived as Amused, Polite, and Embarrassed/Nervous", *J Nonverbal Behav.* (33:1), pp. 17-34

Dhall, A., Goecke, R., Joshi, J., Sikka, K., and Gedeon, T. 2014 "The Second Emotion Recognition In The Wild Challenge", *16th ACM Int'l Conf. on Multimodal Interaction* Istanbul, (EmotiW).

Ekman, P., Davidson, R.J., and Friesen, W.V. 1990. "The Duchenne smile: Emotional expression and brain physiology: II", *J Personality and Social Psychology* (58), pp. 342—353

Engstrom, J., Johansson, E., and Ostlund, J. 2005. "Effects of Visual and Cognitive Load in Real and Simulated Motorway Driving"' *Transportation Research Part F* (8:2), pp. 97—120

Gregor, S., Lin, A.C.H., Gedeon, T., Riaz, A., and Zhu, D. 2014. "Neuroscience and a Nomological Network for the Understanding and Assessment of Emotions in Information Systems Research", *Journal of Management Information Systems*, (30:4), pp. 13-48

Guo, R., Li, S., He, L., Gao, W., Qi, H., and Owens, G. 2013. "Pervasive and unobtrusive emotion sensing for human mental health"' *7th Int'l Conf. on Pervasive Computing Technologies for Healthcare*, pp. 436—439

Hoque, M., Morency, L.P., and Picard, R.W. 2011. "Are You Friendly or Just Polite? Analysis of Smiles in Spontaneous Face-to-Face Interactions", *Affective Computing and Intelligent Interaction, LNCS* (6974) pp. 135—144

Jouni, P., Okko, R., and Serdar, K. 2015 "Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits", *Computer Speech & Language*, (29:1), pp. 145-171 Feature selection code available: http://users.spa.aalto.fi/jpohjala/featureselection/

Kasl, S.V., and Mahl, G.F. 1965. "The relationship of disturbance and hesitations in spontaneous speech to anxiety", *Journal of Personality and Social Psychology*, (1), pp. 425-433

Kraut, R.E. 1978. "Verbal and nonverbal cues in the perception of lying", *Journal of Personality and Social Psychology*, 36:4, pp. 380-391

Lykken, D.T. 1974. "Psychology and the lie detector industry", *American Psychologist*, (29), pp. 725-739

Nakasone, A., Prendinger, H., and Ishizuka, M. 2005. "Emotion Recognition from Electromyography and Skin Conductance", *BSI*, pp. 219—222

Nourbakhsh, N., Wang, Y., Chen, F., and Calvo, R.A.: 2012. "Using Galvanic Skin Response for Cognitive Load Measurement in Arithmetic and Reading Tasks", *24th Australian Computer-Human Interaction Conference*, pp. 420—423

Palanisamy, K., Murugappan, M., and Yaacob, S. 2013. "Multiple Physiological Signal-Based Human Stress Identification Using Non-Linear Classifiers", *Elektronika IR Elektrotechnika*, (19:7), 80-85

Pantic, M., Valstar, M., Rademaker, R., and Maat, L. 2015. "Web-Based Database for Facial Expression Analysis"' *IEEE Int'l Conf. Multimedia and Expo*, p. 5-8

Petridisa, S., Martineza, B., and Pantica, M. 2013. "The MAHNOB Laughter database. Image and Vision Computing", (31:2), pp. 186-202

Picard, R.W., Vyzas, E., and Healey, J. 2001. "Toward machine emotional intelligence: analysis of affective physiological state", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (23:10), pp. 1175-1191

Shangguan, P., Liu, G., and Wen, W. 2014 "The Emotion Recognition Based on GSR Signal by Curve Fitting", *J Information and Computational Science*, (11:8), pp. 2635—2646

Sharma, N., and Gedeon, T. 2014 "Modeling observer stress for typical real environments", *Expert Systems with Applications*, (41:5), pp. 2231—2238

Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. 2012. "A multimodal database for affect recognition and implicit tagging"' *IEEE Transactions on Affective Computing*, (3:1), pp. 42-55

Sun, H., & Zhang, P. 2006. "The role of affect in information systems research: A critical survey and a research model", *Human-computer interaction and management information systems: Foundations*, pp. 295

Valstar, M.F., and Pantic, M. 2010. "Induced Disgust, Happiness and Surprise: an Addition to the MMI Facial Expression Database", *Language Resources and Evaluation*, pp. 65-70

Willis, M. and Jones, C. 2012, "Supporting emotion communication in information systems", 23rd Australasian Conference on Information Systems, Geelong, pp. 1-10

Acknowledgements

The authors would like to thank the participants who took part in this research.