ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/383163009

Modeling the 'Kiss my Ass'-Smile: Appearance and Functions of Smiles in Negative Social Situations

Conference Paper · August 2024

CITATION
1

READS
203

6 authors, including:

Mirella Hladký
Universität des Saarlandes
7 PUBLICATIONS **32** CITATIONS

SEE PROFILE

Xi Laura Cang
University of British Columbia
15 PUBLICATIONS **177** CITATIONS

SEE PROFILE

Karon E. Maclean
University of British Columbia
148 PUBLICATIONS **6,039** CITATIONS

SEE PROFILE

Patrick Gebhard
Deutsches Forschungszentrum für Künstliche Intelligenz
119 PUBLICATIONS **2,694** CITATIONS

SEE PROFILE

All content following this page was uploaded by Mirella Hladký on 17 September 2024.

The user has requested enhancement of the downloaded file.

Modeling the 'Kiss my Ass'-Smile: Appearance and Functions of Smiles in Negative Social Situations

Mirella Hladký*, Rúbia R. Guerra†, X. Laura Cang†, Karon E. MacLean†, Patrick Gebhard*, Tanja Schneeberger*
*German Research Center for Artificial Intelligence, Saarland Informatics Campus, Saarbrücken, Germany,
firstname.lastname@dfki.de
†University of British Columbia, Vancouver, BC, Canada, {rubiarg, cang, maclean}@cs.ubc.ca

***Abstract*—Computational emotion recognition relies on observable expressions. However, negative situations can evoke regulation mechanisms that obscure and mask emotional experiences, often by smiling. As smiles are typically associated with positive emotions, this mismatch of emotional experience and expression may lead to misinterpretations by most current algorithmic affective computing approaches. To improve computational modeling of real-life experiences and expressions in negative social situations, we explore connections between smile appearance and function, incorporating participants' rich personal self-reports into ground truth labels for their expressions. We present an empirically grounded smile corpus of 199 smiles that is based on a) recordings of *N* = 30 participants in negative social situations that are analyzed regarding smile morphology and b) a category system of smile functions based on participants' self-reports. In a computational model, we used cleaned corpus data of 183 unique smile instances to classify five smile function categories based on observable nonverbal signals, with results benchmarked at above chance. Applying a theory- and data-driven approach, our analyses confirm a complex relationship between internal smile functions and observable signals. Finally, we discuss smile functions in negative social situations, including 'despising', 'provoking', and 'kiss my ass'-smiles.**

***Index Terms*—Emotion Modeling, Corpus, Facial Coding**

I. Introduction

By adulthood, we often feel we have the ability to recognize or interpret our own emotional states and those of others. In some cases though, emotional processes go beyond an intuitive understanding: Why do we sometimes laugh when it is entirely unfitting, or smile in an unpleasant social situation? Situations in which facial expressions do not reflect internal emotional states pose a major challenge for computational emotion recognition and modeling. Many approaches rely on observable expressions as mapped to basic emotions described by Ekman in 1993 [1], e.g., [2]–[4]. However, a one-to-one mapping of expression to internal experience does not always reflect human emotional reality [5]: emotions are not always directly observable [6], and may not be experienced consciously. In particular, negative emotions (e.g., shame) tend to be regulated to protect the self and social relations [7]–[9].

For systems that rely on recognition of user emotions, e.g., to enhance social skills [10], [11] or offer therapeutic assistance [12], [13], an accurate understanding of users' internal states is crucial. Systems that neglect the possibility of mismatch between internal experience and observable expressions may react inappropriately and negatively affect users' well-being, particularly when experiencing negative emotions.

Smiles can seem like straightforward indicators of joy, but are, in fact, one of the most complicated emotional expressions [14]. Occurring in both positive and negative contexts, their morphology and connected nonverbal signals vary widely, smiles can serve widely different functions [15], [16]. We know little about the roles smiles play in negative social situations where emotions are often regulated and not directly linked to observable signals [7], [14].

This paper contributes (a) a unique empirically grounded corpus of smiles and their functions in negative social situations, and (b) a computational model benchmarking the connection between smiles' morphological appearance and their self-reported functions. This dataset, model, and evidence of its accuracy support a larger goal of extending emotion recognition from observable cues to internal experience.

II. Background and Related Work

This work is based on a model of emotion that distinguishes *externally communicated* components of emotion, that show in social signals (e.g., smiles), and not directly observable *internal* components [17]. These components may not align as internal experiences do not always reflect in observable signals [5], [6], e.g., a smiling individual may be experiencing negative emotions. The display of emotions depends on emotion regulation strategies to influence emotional experience and expression [8] – consciously or unconsciously. In negative social situations, emerging emotions such as shame are often regulated to protect the self and social relationships [7]–[9].

A. The Social Emotion Shame

Shame occurs when an individual evaluates their actions, feelings, or behavior as not meeting social values, norms, or demands [7], [18], [19]. As such, the experience of shame is often associated with negatively perceived social situations. It is driven by a fear of social rejection from an interaction partner whose opinion is valued [19]. In job interviews, for example, applicants who feel that they do not meet expectations may fear rejection and experience shame. To protect one's self-concept and influence social relations, shame is often regulated (for known shame regulation strategies see [9]) and not openly displayed [8], [17]. Signals of shame and shame regulation include averting one's gaze or head, self-touch of face or body (e.g., [6], [19]), which stem from a wish to hide [7]. Studies of participants' reactions in shame-eliciting situations have found

979-8-3315-1643-7/24/$31.00 ©2024 IEEE

smiles in association with negative social emotions and their regulation [20], [21], contrary to typical expectation.

B. Smiles and Their Functions

Smiles occur in various contexts and serve various functions. Besides joy, they can signal appeasement, but also dominance [15]. Smiles serve social functions by regulating interactions and relationship quality [9], [16]. Deliberate smiles can mask internal emotions in social situations [16]. Smiles can promote or indicate internal emotion regulation [22] by inducing positive emotions [23]; or replacing negative emotions with positive ones [9]. Different smile functions might be linked to different morphological appearances [15]. The minimal requirement to produce a smile is the activation of the lip-corner puller muscle – resulting in a *Non-Duchenne* (also: false, deliberate) smile. *Duchenne* (also: felt, genuine, enjoyment) smiles additionally require activation of muscles surrounding the eyes [16]. Here, we conceptualize laughter (includes vocalizations or respiratory sounds [24]) as a type of smile, as the same muscles are activated. Ekman identified 18 smile types based on morphological aspects [25], including *false*, *embarrassment*, and *contempt* smiles that can occur in negative social situations. Lip corner-puller and eye muscle activation are assumed to distinguish false from felt smiles. Embarrassment smiles are characterized by gaze aversion, pressed lips, and a raised chin [26]. False smiles are often asymmetric or characterized by tightened lip corners and puffed cheeks [24], which include smiles masking true internal emotions. Some smile types result from experiencing multiple emotions simultaneously, e.g., enjoyment paired with contempt can show in tightened lip corners. Certain smile types, such as the listener response smile, serve social functions [25]. As smiles or laughter in negative situations might be viewed as inappropriate, they might be suppressed, resulting in *'smile controls'* (e.g., depressed lip corners, pressed, pursed lips) [26]. Due to their complexity, neither this collection of smile types nor any other is exhaustive.

C. Emotion Recognition and Modeling

Affective Computing as a field seeks to understand user emotions, including recognizing emotions automatically [2]–[4] and in real time [27], [28]. OpenFace – using facial landmark and action unit detection as well as head pose and gaze estimation [29], and MediaPipe Facemesh – a face geometry approximation software [30] – enable the recognition of externally communicated components of emotions. Computational emotion models are presented by [31], [32]. [33] and [34] combine emotion recognition and modeling, which can inform therapeutic assistance [12], [13] and social training systems, with which users can practice social skills and behavioral norms [10]–[12], [35]. These applications rely on accurate interpretation of emotional expression – trained to consider smiles simply as signals of joy, they may fail with negative consequences when the true experience is negative. A deeper approach to interpret internal emotion requires understanding the connection between morphological appearance of smiles and their functions in negative social situations.

III. Smile Corpus

This work investigates how different morphological signals of smiles are linked to different smile functions in negative social situations. We collected a corpus of smiles in negative social situations. To create negative social situations, we applied a validated and controlled protocol to elicit the social emotion shame [20], [21]: in an online job interview with a virtual agent, we confronted participants with three shame-eliciting situations. The corpus contains multi-modal data on smile morphology and smile functions from behavioral observations of participants' nonverbal behavior, and semi-structured interviews on smile functions. The data collection protocol was approved by the Ethical Review Board of Saarland University. The data collection and validation (in-depth interviews and qualitative content analyses of interview and observational data) were conducted by a trained psychologist.

A. Participants

The smile corpus consists of data from $N = 30$ participants (24 female; 6 male; aged 19-58 years with *M* = 30.33, *SD* = 10.59; 11 in employment). The sample size was determined by recommendations for qualitative in-depth data collection and analysis [36]. All had least one job interview before (*M* = 6.3, *SD* = 4.8). For two reasons critical to successful data collection, we recruited psychology students: 1) they have above-average abilities to access and reflect on internal experiences (critical for self-report collection) [37], and 2) are still exploring careers and can likely imagine themselves in a job interview (critical for role play immersion). They were rewarded with course credit. Overall, we collected video data from 90 shame-eliciting situations and 30 post-interviews referring to these situations.

B. Data Collection Procedure

1) Introduction: Via video chat, a human experimenter welcomed participants and explained the study procedure including an online job interview with an autonomous virtual agent with a female human-like appearance [38]. A picture of the agent can be found on OSF[1]. Participants were asked to imagine themselves in a job interview for a much desired position. They completed a questionnaire that included consent, demographics and items assessing their current affective state.

2) Shame-eliciting Job Interview: The experimenter then explained that the virtual agent would take over for the job interview during which the participant should interact human-like (e.g., natural language, nonverbal behavior). The agent was described as autonomous; however, in reality the experimenter controlled the timing of the agent's pre-scripted behavior. Throughout the job interview, which took on average 5 minutes, the virtual interviewer made three shame-eliciting statements and gave participants time to react to them: 1) *Before we start, a quick question. Where did you get that outfit? Somehow it doesn't really suit you.* 2) *What you said has already been said by all the other applicants. You didn't*

[1]https://osf.io/dm6jx/?view_only=b13d7edaeed54cfa888ccf7a417e605f

exactly stand out. 3) *Well, that answer was not very impressive. I've heard better from the other applicants.*

After the job interview, participants assessed their current affective state again and learned that the study aimed to investigate how individuals reacted to and coped with unpleasant situations. For the complete job interview script visit OSF[1] .

3) Semi-structured Post-Interview: The subsequent post-interview (35 minutes on average) addressed experienced affect, and functions of any smiles that occurred during the shame-eliciting situations. Participant and experimenter together watched a video of the three shame-eliciting situations. First, participants self-reported on their experienced affect in each situation. Then, the video was stopped and re-watched at every smile instance shown during the situation and participants were asked to remember and verbalize why they smiled in the specific instance. Individual interview variations and open-format questions allowed for an in-depth exploration of participants' internal experiences. Examples are: *What do you think, why did you smile here?*; *Did you want to achieve or show something specific with this smile?*; *Did the smile help you in any way to deal with the situation?* For every smile instance, one or multiple functions were described. We followed guidelines by [21] and [39] and applied techniques to establish positive rapport between experimenter and participant [21]. Finally, participants completed a post-interview evaluation, were debriefed and rewarded. The full data collection procedure took about one hour.

C. Validation of the Data Collection Procedure

Three steps were taken to validate the data collection procedure. The first two steps validated that the shame elicitation protocol by [21] successfully created negative social situations.

1) Using the *State Shame and Guilt Scale* [40], shame was self-reported as significantly ($t(29) = 2.69$, $p = .012$, $d = 0.64$) higher after ($M = 1.77$, $SD = 0.64$) than before the job interview ($M = 1.43$, $SD = 0.39$) – a medium to large effect [41]. Distractor items were included to avoid priming.

2) Self-reports of experienced affect during the shame-eliciting situations were analyzed according to the *qualitative content analysis (QCA)* protocol *inductive category formation* by [42] in the software QCAmap[2]. A trained rater formed categories based on 236 affect descriptions. Here, 'affect' means all internal experiences. 66% were categorized as shame or related negative social affect (e.g., unpleasant, attacked, inferior, embarrassed, insecure). 18% of descriptions reflected known shame regulation strategies [9]. In 11% of descriptions, participants reported only slight or no experienced shame and, in 5%, positive affect (e.g., self-confident, amused). Often, multiple affects were described for one situation.

3) In a 4-item *post-interview evaluation* by [21] (5-point Likert scale; 1 = strongly disagree, 5 = strongly agree), participants rated their candor in the post-interview and its agreeableness as high ($M = 4.5$, $SD = 0.37$, $min = 3.5$) – see [21] for items. This was supported by free comments: All were positive (e.g., "Self-reflecting was interesting and fun").

[2]www.qcamap.org

D. Collected Data

The smile corpus consists of multi-modal data on two variables: smile morphology and smile functions.

1) Smile Morphology: We defined an *annotation scheme* based on existing research and theories about nonverbal signals associated with smiles in negative social situations (see II) using the Facial Action Coding System [43] as a reference. It includes the following labels: **Duchenne** (smile involving: lip corner puller, eye muscle); **Non-Duchenne** (smile involving lip corner puller only); **Non-Symmetric** (unilateral or uneven smile); **Intensity** (smile with teeth showing); **Laughter** (audible vocalization or respiration); **Smile Control** (lip actions that control or suppress a smile: Lip Corner Depressor, Chin Raiser, Purse Lip, Press Lip, Suck Lip, Cheek Puffer, Lip Corner Tightener); **Gaze Aversion**; **Adaptors** (body/face self-touch); **Situation** (defines start and end of the shame-eliciting situations, i.e. video sections considered for analysis).

A trained rater manually applied the annotation scheme to the video material of shame-eliciting situations using the annotation tool NOVA [44], following the standardized QCA protocol *deductive category assignment* by [42]. Visit OSF[1] for the complete annotation scheme and example screenshots of annotations in NOVA.

2) Smile Functions: Self-reports about 199 smile instances to which participants assigned a total of 364 functions (multiple functions could be assigned to one smile instance) were obtained in semi-structured post-interviews (see III-B).

A trained rater analyzed the self-reports according to the standardized QCA protocol *inductive category formation* by [42] in QCAmap[2]: Descriptions of smile functions in the transcribed post-interviews were systematically labelled, summarized into initial functions and structured into higher-level categories. Table I shows the resulting category system with example quotes of participants and category frequencies.

In a first step, 28 initial functions were inductively extracted from participants' self-reports. Further categorization led to three main categories: *Representative Functions* (1) refer to smiles that represent an emotional experience. *Intrapersonal Functions* (2) include smiles as manifestations or triggers of internal processes, such as emotion regulation. *Interpersonal Functions* (3) represent smiles related to social relationship regulation. This category accounts for the largest amount (65%) of all functions. It has three distinct sub-categories: *Relationship Support* (3.1.) is further sub-divided in three sub-categories related to enhancing or maintaining the relationship. *Relationship Adjustment* (3.2.) represents smiles that are related to intentions of abandoning the relationship or adjusting one's social status within it. *Mask Internal Emotions* (3.3.) represents smiles that aim to cover up true internal emotions.

As recommended by [42], all performed QCAs were validated via intra-rater-agreement. Finally, we linked the morphological aspects of each smile instance to the functions registered for the particular instance. This data was used for building a machine learning model.

Table I: Category system of smile functions as result of qualitative content analysis.

Main functions and their sub functions			Initial Functions	Participants' example quotes
1. Representative Functions (19%)			insecurity	'I didn't know what to say', 'Because I had not expected it at all and was very insecure'
			anger	'I'm super annoyed with her and that's why a pissed pressed smile', 'I was furious'
			shame/embarrassment	'I smiled because I was embarrassed', 'I was bit ashamed as I couldn't think of anything to say'
			unspecified	'I don't remember if it was because I didn't feel taken seriously or due to negative emotions'
			amusement	'Because I found the situation somehow absurdly funny', 'comedy of the situation'
2. Intrapersonal Functions (16%)			enhance well-being	'Smiling makes you feel better', 'It was to calm myself down'
			internal self-regulation	'It was rather regulating and totally spontaneous. It helped me not to escalate', 'The short smile was more for myself. I grinned briefly and quickly tried to regulate myself'
			self-satisfaction	'I am satisfied with what I have said, and I smile', 'I expected this negative statement and smiled because I was right'
			self-reflection	'I was underlining what she was criticizing', 'I thought: What did I say? That wasn't very polite.'
3. Interpersonal Functions (65%)	3.1. Relationship Support (27%)	3.1.1. General Relationship Enhancement	appeasement	'I smiled to relax the situation', 'I didn't want confrontation, because I wanted the job'
			enhance relationship	'I still have to communicate with the person and smiling always helps quite well'
			display acceptance of other's opinion	'I want to show her 'okay, this is new for me, but I gladly accept the criticism', 'With such a smile you also signal that you accept what the other person says'
			keep pos. atmosphere	'Suspect that I did that to keep the mood from tipping', 'Simply for a positive atmosphere'
		3.1.2. Social Display Rule	show courtesy	'I squeezed out a smile to stay polite', 'it's still a job application situation and you smile anyway'
			disagree in socially accepted manner	'I contradicted her but in a socially acceptable way', 'By still being able to smile joyfully, I show that I can't quite agree with her'
			self-control / avoid slip-through	'That was a last attempt to stay friendly', 'I pulled myself together so I wouldn't tell her what I thought of her and the conversation'
		3.1.3. Positive Social Self-Representation	display ability to deal with negative situation	'I show that I can laugh about it, so that I look good in front of her', 'I wanted to show her that even in unpleasant situations where I am attacked, I can keep a polite facial expression'
			display professionalism/ seriousness	'To restraint from saying anything else and remain professional towards the interviewer', 'I had to smile but I actually tried to say it seriously, so I tried to suppress it'
			display self-confidence	'I want to demonstrate to her that I won't be influenced by her negative feedback', 'I was so self-confident in that moment'
			pos. self-presentation	'to appear sympathetic', 'to look good in front of her'
	3.2. Relationship Adjustment (27%)		show dominance / status	'I put myself very much above her', 'to show superiority', 'to gain back the control'
			resignation	'There's nothing I can say anymore anyway', 'Resignation and waiting to see what follows'
			depreciate other	'I made fun of her and ridiculed the situation', 'That's a pretty despising smile and I also had contemptuous thoughts'
			reject other	'This 'kiss my ass'-smile appeared', 'Because I wanted nothing more to do with her'
			provoke/attack other	'It was a provoking smile', 'The best way to show your teeth to your opponent is to smile'
			highlight other's faux pas	'To show that it attacked and hurt me, I suppressed the smile', 'I realized I am not taken seriously and am criticized non-stop, so I suppressed the smile'
	3.3. Mask Internal Emotions (10%)		mask discomfort	'I was a bit embarrassed, but I just tried to put on a smile', 'I wanted to hide the discomfort'
			mask unspecified feelings	'I kind of show it but not entirely', 'It's like a wall, so that I still look friendly to the outside'

Note: In brackets: Percentage share in the total number of functions for each main function and level 1 sub-functions of Interpersonal functions (3). The column left of the example quotes contains the 28 initial categories extracted directly from participants' descriptions forming the basis for the higher level categories. Participants often described multiple functions for one smile instance.

IV. Machine Learning Model

Modeling relationships between morphological signals of smiles (features) and functions they serve (class targets) takes us from simple observation to identifying complex interaction patterns. We outline our machine learning pipeline, data preprocessing, model selection and evaluation, and the significance of specific features. Figure 1 provides an overview.

A. Data Preprocessing

We extracted relevant features directly from the input data, using all pre-defined smile signals (III-D and Table I).

1) Choice of Function Level: We chose the level at which to include smile functions as ground truth classes such that function categories are represented in proportion to the amount of data they represent. *Representative (1)* and *Intrapersonal Functions (2)* were labelled at the main level as they represent a smaller share (19%, 16%) of all functions and do not have sub-categories. Interpersonal Functions (3) represent the largest data amount (65%) and subsume conceptually distinct sub-categories. Thus, the sub-categories *Relationship Support* (3.1), *Relationship Adjustment* (3.2) and *Mask Internal Emotions* (3.3) are included as labels. In total, five functions act as ground truth class labels covering all identified smile functions (Table I). As one data instance can have multiple functions, each function is modelled independently using a binary classifier to label presence or absence.

2) Final Composition: From the corpus of 199 smiles, we removed 'duplicate' data instances where signals and function were identical to avoid bias due to over-representation of any one signal combination. All remaining 183 data instances used for model building are unique. All models are built on 183 unique data instances: **61** instances of Representative Functions, **54** Intrapersonal Regulation, **78** Relationship Support, and **63** Relationship Adjustment. **38** Mask Internal Emotions.

3) Class Balancing: At the most challenging case, chance for binary classification varies from 20.76% for Mask Internal Emotions to 42.62% for Relationship Support; therefore, we report both % accuracy and micro F1-scores in Figure 2. Using scikit-learn [45], we perform class balancing by setting the class-weight hyper-parameter during smile function classification. Classes are weighted to be inversely proportional to their frequencies with smaller classes being given more weight during training. All features are converted to binary to represent presence or absence (0 or 1).

B. Model Selection

We constructed a model of multiple binary classifiers to recognize each instance's smile function – effectively asking "Is this smile's function to Mask Internal Emotions?" for each function. To determine optimal hyperparameters by classifier, we used grid search with stratified 5-fold cross-validation, systematically evaluating a range of configurations to select the best performing combination on validation data. We explored a diverse set of commonly used classifiers, including Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbors, Random Forest, Gaussian Naive Bayes, Support Vector Machine, and XGBoost. After grid search, we refitted the best classifier for each smile function using 30 random train-test splits (80%/20%), which helps to assess generalizability and robustness to data variation. Each binary classifier was evaluated using accuracy and micro F1 scores as primary metrics.

C. Model Evaluation

We evaluated trained binary classifiers, examining how combinations of smile signals may relate to functions. Runtime for all experiments takes around 20min on a machine with Apple M1 Max (64GB).

1) Performance: For each smile function, we recorded scores from the best classifier as determined by grid search with stratified 5-fold cross-validation. For test accuracy measures (see Figure 2), all smile function classifications exceed chance (50% for presence or absence of function in a balanced scenario). We note that smiles of Mask Internal Emotions were the best performing at ∼82% accuracy and with the highest F1-scores (however, this class has the lowest data count so we investigate further below, see IV-D) and Relationship Support exceeds but is nearest chance with median accuracy at 57%.

2) Dimensionality Reduction: We conducted a principal component analysis (PCA) on the original morphological signals. To explore how dimensionality reduction of the feature set influences classification, we compared how the top components performed as features against the original set. Dimensionality reduction may improve performance for smile functions (Figure 2). The Smile Control signal contributes most prominently for Representative Functions with or without dimensionality reduction (see Figures 2 and 3). Figure 2 relates the components to the original morphological signal features. Here, we report the top three components where the component capturing the most variability – PC1 – is most influenced by whether the signal is Duchenne or not, and Intensity. PC2 is dominated by Gaze Aversion and Non-Symmetry, while PC3 heavily relies on Smile Control.

D. Feature Importance Analysis

To reveal how each feature contributes to the model's predictions, we employed SHAP (SHapley Additive exPlanations) [46]. We computed the importance of each smile signal feature (S) by calculating mean model performance when the feature is vs. is not included. Figure 3 shows absolute SHAP values for smile signals (S) in predicting smile functions (F) that rank each feature's impact on the model's output (0-1 = no to high importance) – e.g., Non-Duchenne is highly influential for Relationship Support. Results for Mask Internal Emotions suggest that the model is incapable of achieving an interpretation and predicts according to class biases, challenging the high performance accuracy.

V. Discussion

This paper presents a benchmark of a computational model of morphological appearance and functions of smiles in negative social situations. The basis of the model is a unique smile

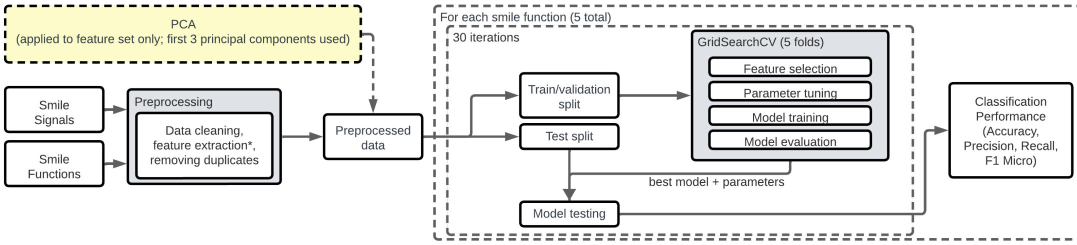

Figure 1: Classification pipeline. Preprocessing generated 183 unique data instances of Smile Signal features with binary Smile Function labels (presence or absence), we report mean classification performance over 30 test iterations for each of the five Smile Functions, trained separately by function. We also report classification results after dimensionality reduction (using PCA) on the feature set. As multiple functions may relate to one smile, classes are not mutually exclusive. For each iteration, we use the training data to determine the best combination of model and parameters (for that split) using Grid Search Cross Validation (5 folds). After determining the best model, we refit the model using the training data and test it using the test split, from which classification performance metrics are derived (accuracy, recall, precision, F1 micro score).

corpus with multi-modal data from behavioral observations of participants in negative social situations and interviews about internal experiences and displayed smiles in these situations. Our results indicate that we have successfully created negative social situations using the shame elicitation protocol by [21].

A. *Smile Functions*

Based on participants' self-reports, we developed an extensive categorization of smile functions in negative social situations, featuring three main categories. Most smiles served *Interpersonal Functions*, which included not only smiles that support and positively influence the relationship (e.g., by signaling appeasement, courtesy), consistent with [15]. Remarkably, an equal share of Interpersonal smiles indicated conflict, dominance (supporting [15]), devaluation, or even relationship abandonment, e.g., 'provoking', 'despising' and 'kiss my ass'-smiles. A few Interpersonal smiles served to mask internal emotions, as posited by [16]. We also found smiles directly representing negative emotions (*Representative Functions*), e.g., 'embarrassed', 'insecure', and 'pissed' smiles. Smiles with *Intrapersonal Functions* served to regulate emotions and enhance well-being or self-image (supporting [22], [23]). Our data shows that smiles represent negative internal emotions and regulate relationships and emotions – highlighting the importance of differentiating between externally communicated and internal components of emotions [17].

B. *Interrelation of Smile Functions and Morphology*

To identify patterns between functions and morphological appearance of smiles, we generated a machine learning model of five classifiers for Representative Functions, Intrapersonal Functions and for the three sub-categories of Interpersonal Functions: Relationship Support, Relationship Adjustment and Mask Internal Emotions. Based on the connected morphological signals, we present a classification of all five smile functions at better than chance performance. The 30-person corpus was an arduous collection and labelling procedure that generated 199 smiles (183 unique), including self-reported functions, with low repetition due to complex individual experiences and expressions. This highly variable and relatively small dataset presents a challenge for machine classification, precluding the use of deep learning. Under these constraints, we present a benchmark for the classification and focus on connecting feature contribution to established theory.

Our analyses demonstrate the importance of morphological signals in predicting individual functions. Differentiating between Non-Duchenne and Duchenne smiles on a morphological level is important for predicting several smile func-

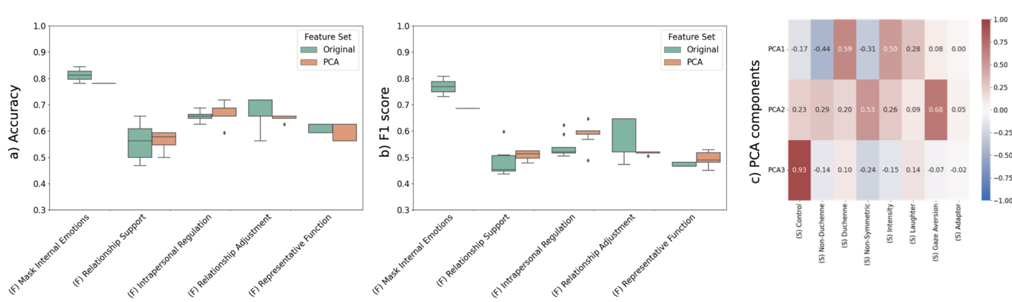

Figure 2: Box-and-Whisker Plots of the Performance of Smile Function Models by Percent Accuracy (a) and F1-score (b). The box plots show 1.5 IQR distribution of scores across 30 random train-test splits on the best performing classifier (determined by grid search) by smile function. By predicting the presence or absence of a smile function, balanced-chance accuracy is 50% (0.5). High F1-scores high precision and recall, indicating fewer false positives or false negatives. (c) Composition of principal components using original signals. High magnitudes represent strong influence of a particular signal for that component.

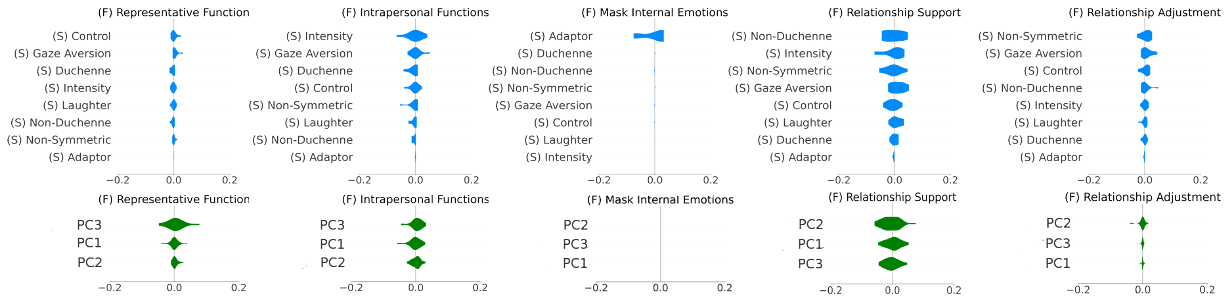

Figure 3: Feature importance (S) for smile function prediction (F) using absolute SHAP values. SHAP values, ranging from -1 to 1, represent the contribution of each feature to the model's prediction for a given sample, compared to baseline prediction. A value of -1 decreases the likelihood of prediction, 1 increases it, 0 has no impact. If no negative values are observed, it suggests that the features consistently increase the prediction. Features are sorted according to their importance for each function.

tions. However, the interpretation of Non-Duchenne smiles as false or deliberate and Duchenne as felt or genuine enjoyment smiles should be viewed critically. Interestingly, in our model, Duchenne and intense smiles represent negative internal emotions (e.g., shame, insecurity) and are connected to emotion regulation processes, e.g., enhancing well-being (Intrapersonal Functions). That is, those smiles help to cope with negative social situations, adding to existing findings [9], [22], [23]. Non-Symmetric smiles are connected to Relationship Adjustment, but also Intrapersonal and Relationship Support functions, although other researchers interpreted them as false [24], contempt or masking smiles [25]. Smile Controls are influential for Relationship Adjustment, but also for Relationship Support functions, which is remarkable as these represent opposite goals of maintaining or enhancing versus compromising or abandoning a relationship. So far, Smile Controls were known as means of suppressing socially inappropriate smiles [26]. Adaptors (body/face self-touch) is the only feature that classifies Mask Internal Emotions (while unrelated to other functions), suggesting that self-touch serves to hide from another's gaze [7] and to hide emotions. Note that this function had the lowest instance count, warranting further investigation for stronger conclusions.

In some cases, signal *combinations* show stronger predictive power. For instance, to predict Relationship Support, a combination of all signals (except Adaptor) is required. Smile Control combined with, e.g., a Non-Symmetric smile, is a strong predictor for Representative Functions. It is intuitive that smiles representing negative internal emotions are controlled and altered instead of natural and open. A combination of Gaze Aversion and Non-Symmetry in smiles indicates an intention to compromise or abandon a relationship (Relationship Adjustment) – Non-Symmetry is a known indicator for false smiles [24] and breaking eye contact may indicate rejection. Overall, our data supports that a simple one-to-one mapping of expressions to internal emotions is not sensible [5].

We designed our model to be transparent and reveal in what way internal experiences are connected to observable signals, such that it can inform psychological theories and models. A strength of our model lies within the combination of theory- and data-driven approaches. It is informed both by theoretical knowledge about nonverbal behavior in negative social situations and by empirical data that directly reflect participants' reports about their internal experiences.

C. Limitations and Future Work

To provide an initial benchmark for this novel corpus and model, we focused on functions and associated morphology of smiles in the context of negative social shame-eliciting situations in a limited sample. To optimize accuracy and generalizability, future research should explore other contexts and populations, and include further smile variables like duration, onset, offset, and apex. Smile morphology and function analyses should be re-verified by a second (FACS trained) rater. Classification rates may improve when optimizing for performance metrics – here, we have structured our model to stay true to participants' experiences and theoretical grounding. Our data collection procedure is limited by the challenge of verbalizing internal experiences in unpleasant situations, to which emotion regulation processes may restrict access [8], [47]. Our results support that internal experiences, and whether and how they reflect in observable expressions, vary between individuals, making it difficult to identify clear patterns and relationships in the data [5]. We propose that individual models may reflect the reality of human emotions more accurately.

VI. Conclusion

This paper presents a novel computational model of smile functions and their morphological appearance in negative social situations. Combining a theory- and data-driven approach, our model classifies smile functions based on nonverbal signals, with results benchmarked at above chance, confirming a complex relationship. Our model is based on a unique empirically grounded multi-modal smile corpus of 199 smiles – which we make available for research purposes in anonymized form on OSF[1] – consisting of annotated video data of morphological appearance of participants' smiles and a corresponding category system of self-reported functions that smiles served in negative social situations. We contribute to the effort to build emotion models and recognition systems that consider not only externally observable but also internal components of emotions and highlight that their one-to-one mapping is not feasible.

Ethical Impact Statement

All captured data is treated confidentially, stored securely restricted to authorized personnel and used for scientific purposes solely. The collected data included personally identifiable information such as facial videos and interview recordings. All participants were informed and consented to the data collection and the use of this data for our research purposes. They consented to the publication of their data in an anonymized form, after the removal of personally identifiable information. We make our corpus available only in this anonymized, processed form. Participants were informed about their right to withdraw their consent. They were aware that their participation was voluntary and that they could abort the study at any time – they would still receive their course credit reward. The data collection protocol was approved by the Ethical Review Board of Saarland University.

During our study, participants experienced shame-eliciting social situations during a job interview role-play with a virtual agent. Participants were made to believe that they were alone with the virtual agent during the job interview and that it interacted automatically and adaptively, while the experimenter discreetly observed the situation and controlled the agent's pre-scripted behavior to create the impression of a natural interaction. As an inherent part of this study was to elicit emotions, this setup was important for the job interview to be immersive and realistic. Also, the post-interview about participants' internal affect and functions that their smiles served during the experienced shame-eliciting situations, was important for gathering data for our corpus and model. We are aware that experiencing situations that elicit unpleasant emotions and elaborating on their internal experiences during those situations with the experimenter can be challenging for participants. The experimenter was a trained psychologist that applied techniques to counterbalance those challenges. Techniques to establish positive rapport between experimenter and participant suggested by [21] were applied. Throughout the study, the experimenter asked participants how they were feeling and comforted them with empathetic responses. Overall, the experimenter took measures to create a friendly and warm atmosphere and ensured the participants' well-being. Participants were fully debriefed at the end, also about the virtual agent's pre-scripted behavior and dialogue which was not giving individualized responses to what participants had said. The experimenter released the participants after assessing that the emotion-eliciting experimental procedure and the post-interview had no persisting negative effects on their well-being. Regarding the post-interviews, we received positive feedback from participants stating that it was, for example, agreeable, interesting, and fun.

Our corpus is based on a sample of 30 psychology students that are assumed to have higher awareness of and better access to their internal experiences [37]. While this sample is useful to train our model due to high data validity, the data collection process may be biased by it and its replicability might be limited when applying it to a different or more diverse sample. Our participants had similar cultural backgrounds, so the data might not represent other cultures and demographic backgrounds. As mentioned above, we assume that internal experiences and nonverbal expressions, as well as their connection, differ between individuals. Our model is a first step toward understanding this connection. However, we suggest that individual models may more accurately represent and capture the reality of human emotion.

Our model contributes to efforts of interpreting internal states. While is has benefits for Affective Computing systems that rely on an understanding of the user's emotions, such as social training systems or therapeutical assistants, a potential misuse of applications raises privacy concerns. In our study, we asked participants about what functions their smiles served in negative social situations. While they consented to it for research purposes, in real life situations, people might not want functions of their smiles and thus internal experiences and underlying intentions to be revealed, which might have negative consequences for social interactions and relationships. It is essential for users to be informed about and consent to the functionality, data collection and processing and potentially associated risks, before using systems that apply models to interpret observable expressions and internal states. Such systems should not be used without the knowledge and consent of people that might be observed by the system. An unconsented application can have negative consequences for observed individuals as deeply private information about their internal experiences might be involuntarily collected, exposed and used in ways that could damage the individual's social reputation, privacy and well-being.

Acknowledgment

This work is funded by the German Ministry of Education and Research (BMBF) within the UBIDENZ project (Funding Code 13GW0568D).

References

[1] P. Ekman, "Facial expression and emotion." *American psychologist*, vol. 48, no. 4, p. 384, 1993.

[2] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *Trans. on Pattern Analysis & Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.

[3] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M. Torres, S. Scherer, G. Stratou, R. Cowie, and M. Pantic, "Avec 2016: Depression, mood, and emotion recognition workshop and challenge," in *Int. Workshop on Audio/Visual Emotion Challenge*, 2016, pp. 3–10.

[4] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 211–223, 2012.

[5] L. Feldman Barrett, *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt, 2017.

[6] D. Keltner, "Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame." *Journal of personality and social psychology*, vol. 68, no. 3, p. 441, 1995.

[7] M. Lewis, "Self-conscious emotions: Embarrassment, pride, shame, and guilt." in *Handbook of Emotions*, M. Lewis, J. M. Haviland-Jones, and L. Feldman Barrett, Eds. Guilford, 2008, vol. 3, p. 742–756.

[8] J. J. Gross, *Handbook of emotion regulation*. Guilford, 2013.

[9] D. L. Nathanson, *Shame and pride*. Norton, 1994.

[10] P. Gebhard, T. Schneeberger, E. André, T. Baur, I. Damian, G. Mehlmann, C. König, and M. Langer, "Serious games for training social skills in job interviews," *IEEE Transactions on Games*, 2018.

[11] T. Schneeberger, N. Sauerwein, M. S. Anglet, and P. Gebhard, "Stress management training using biofeedback guided by social agents," in *26th Conference on Intelligent User Interfaces*, 2021, pp. 564–574.

[12] P. Gebhard, T. Schneeberger, M. Dietz, E. André, and N. u. H. Bajwa, "Designing a mobile social and vocational reintegration assistant for burn-out outpatient treatment," in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 2019, pp. 13–15.

[13] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet *et al.*, "Simsensei kiosk: A virtual human interviewer for healthcare decision support," in *2014 Conference on Autonomous Agents and Multi-Agent Systems*, 2014, pp. 1061–1068.

[14] P. Ekman, W. V. Friesen, and M. O'Sullivan, "Smiles when lying," *What the Face Reveals. Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, pp. 201–214, 1997.

[15] U. Hess, M. G. Beaupré, N. Cheung *et al.*, "Who to whom and why–cultural differences and similarities in the function of smiles," *An empirical reflection on the smile*, vol. 4, p. 187, 2002.

[16] P. Ekman and W. V. Friesen, "Felt, false, and miserable smiles," *Journal of nonverbal behavior*, vol. 6, no. 4, pp. 238–252, 1982.

[17] U. Moser and I. Von Zeppelin, "Die Entwicklung des Affektsystems," *Psyche*, vol. 50, no. 1, pp. 32–84, 1996.

[18] J. J. Gross, O. P. John, and J. M. Richards, "The dissociation of emotion expression from emotion experience," *Personality and Social Psychology Bulletin*, vol. 26, no. 6, pp. 712–726, 2000.

[19] W. K. Hahn, "The experience of shame in psychotherapy supervision." *Psychotherapy*, vol. 38, no. 3, p. 272, 2001.

[20] T. Schneeberger, M. Scholtes, B. Hilpert, M. Langer, and P. Gebhard, "Can social agents elicit shame as humans do?" in *Conference on Affective Computing and Intelligent Interaction*, 2019, pp. 164–170.

[21] T. Schneeberger, M. Hladkỳ, A.-K. Thurner, J. Volkert, A. Heimerl, T. Baur, E. André, and P. Gebhard, "The deep method: Towards computational modeling of the social emotion shame driven by theory, introspection, and social signals," *IEEE Transactions on Affective Computing*, pp. 1–16, 2023.

[22] B. L. Fredrickson and R. W. Levenson, "Positive emotions speed recovery from the cardiovascular sequelae of negative emotions," *Cognition & emotion*, vol. 12, no. 2, pp. 191–220, 1998.

[23] C. C. Neuhoff and C. Schaefer, "Effects of laughing, smiling, and howling on mood," *Psychological reports*, vol. 91, pp. 1079–1080, 2002.

[24] W. Ruch, "Extraversion, alcohol, and enjoyment," in *What the face reveals*, P. Ekman and E. L. Rosenberg, Eds. Oxford, 1997.

[25] P. Ekman and S. Wiltschek, *Weshalb Lügen kurze Beine haben: Über Täuschungen und deren Aufdeckung im privaten und öffentlichen Leben*. De Gruyter, 1989.

[26] 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," *Nonverbal behavior*, vol. 33, no. 1, pp. 17–34, 2009.

[27] T. Baltrušaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic action unit detection," in *2015 11th Conference on Automatic Face and Gesture Recognition*, vol. 6, 2015, pp. 1–6.

[28] Z. Zhang, S. Zhai, and L. Yin, "Identity-based adversarial training of deep cnns for facial action unit recognition." in *BMVC*, 2018, p. 226.

[29] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 59–66.

[30] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time facial surface geometry from monocular video on mobile gpus," *arXiv:1907.06724*, 2019.

[31] S. C. Marsella, J. Gratch, and P. Petta, "Computational models of emotion," in *Blueprint for Affective Computing*, K. R. Scherer, T. Bänzinger, and E. B. Roesch, Eds. Oxford, 2010, pp. 21–41.

[32] C. Conati and H. Maclaren, "Empirically building and evaluating a probabilistic model of user affect," *User Modeling and User-Adapted Interaction*, vol. 19, no. 3, pp. 267–303, 2009.

[33] M. Belkaid and N. Sabouret, "A logical model of ToM for virtual agents in the context of job interview simulation," *arXiv:1402.5043*, 2014.

[34] P. Gebhard, T. Schneeberger, T. Baur, and E. André, "Marssi: Model of appraisal, regulation, and social signal interpretation," in *Int. Conference on Autonomous Agents and MultiAgent Systems*, 2018, pp. 497–506.

[35] M. E. Hoque, M. Courgeon, J.-C. Martin, B. Mutlu, and R. W. Picard, "Mach: My automated conversation coach," in *2013 Joint Conference on Pervasive and Ubiquitous Computing*, 2013, pp. 697–706.

[36] B. Fridlund and C. Hildingh, "Health and qualitative analysis methods," *Qualitative research methods in the service of health*, pp. 13–25, 2000.

[37] J. A. Hall, "Psychological-mindedness: A conceptual model," *American Journal of Psychotherapy*, vol. 46, no. 1, pp. 131–140, 1992.

[38] T. Schneeberger, P. Gebhard, T. Baur, and E. André, "Parley: A transparent virtual social agent training interface," in *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*. Association for Computing Machinery, 2019, p. 35–36.

[39] A. Galletta, *Mastering the semi-structured interview and beyond: From research design to analysis and publication*. NYU press, 2013, vol. 18.

[40] D. Marschall, J. Saftner, and J. P. Tangney, *The state shame and guilt scale*. Fairfax, 1994.

[41] J. Cohen, "Statistical power analysis for the behavioral sciences," 1988.

[42] P. Mayring, *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. AUT, 2014.

[43] P. Ekman and W. V. Friesen, "Measuring facial movement," *Environmental psychology and nonverbal behavior*, vol. 1, pp. 56–75, 1976.

[44] T. Baur, I. Damian, F. Lingenfelser, J. Wagner, and E. André, "NovA: Automated analysis of nonverbal signals in social interactions," in *Human Behavior Understanding*. Springer, 2013, pp. 160–171.

[45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[46] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[47] L. Feldman Barrett, "Feelings or words? understanding the content in self-report ratings of experienced emotion," *Journal of Personality and Social Psychology*, vol. 87, no. 2, pp. 266–281, 2004.

View publication stats