A Computational Study of Cultural Effects on Facial Expressiveness

Rafah Ali
BS Computer Science,
Northwestern University

Eshan Tarneja
BS/MS Computer Science,
Northwestern University

Ikhlas Attarwala
MS Artificial Intelligence,
Northwestern University

ABSTRACT

Affect analysis using computational software has attracted great interest in the affective computing community, particularly the recognition of action units and their correlation to emotions. There are still several unexplored areas with regards to representation of different ethnic groups and evaluation of computational tools. In this study, we review OpenFace action unit ratings as compared to ratings by humans of two groups (Caucasians and South Asians) to shed light on this fundamental question. We take a special interest in South Asians to attempt to justify a South Asian category (this is due to South Asians typically being merged into a broader "Asian" category despite evident differences between East Asians and South Asians). We analyze these solutions by comparing computational software to human ratings (split into in-group and out-group categories) of expressed anger. We discuss trends in action units found on South Asian and Caucasian faces and differentiating action units found in our study. We identify trends in the ways that computational software and individuals rate faces across categories. This study allows us to identify possible trends between racial groups, and to define future directions for affect recognition systems.

KEYWORDS

Facial Analysis, Affective Computing, Cultural Differences, OpenFace, Computational Study

1 Introduction & Motivation

We chose to analyze South Asians for two reasons: (1) There has been a large, varied perception on whether or not South Asians fall into the category titled "Asian" or "Asian American" alongside Eastern Asians, an ethnic group that has remarkably distinctive features from South Asians; (2) South Asians have also been largely underrepresented as a group in facial expression research and collection (CAS-PEAL, HKPolyU, CUHK, TFEID). For context, the term "Eastern Asians" can classify individuals with origins from, but not limited to: China, Japan, Taiwan, and Hong Kong. The term "South Asians" can classify individuals with origins from, but again not limited to, India, Pakistan, Afghanistan, Nepal, and Bhutan. Clearly, there is a rich diversity within these two categories, and so we imagine that generalizing across the term "Asian" while addressing the ethnic groups within can lead to unintentional algorithmic bias. We chose not to focus on Eastern Asians for this experiment, as there can be a large origin overlap between both races that we may not have been able to account for. We hoped to elucidate the merits of South Asians independently, and so we chose to compare them to a group that many algorithms are optimized for [5], and who have lesser potential of cultural or origin overlap with South Asians. This study compares perception and facial expressiveness between South Asians and Caucasians.

We chose to analyze South Asians for two reasons: (1) There has been a large, varied perception on whether or not South Asians fall into the category titled "Asian" or "Asian American" alongside Eastern Asians, an ethnic group that has remarkably distinctive features from South Asians; (2) South Asians have also been largely underrepresented as a group in facial expression research and collection (CAS-PEAL, HKPolyU, CUHK, TFEID). For context, the term "Eastern Asians" can classify individuals with origins from, but not limited to: China, Japan, Taiwan, and Hong Kong. The term "South Asians" can classify individuals with origins from, but again not limited to, India, Pakistan, Afghanistan, Nepal, and Bhutan. Clearly, there is a rich diversity within these two categories, and so we imagine that generalizing across the term "Asian" while addressing the ethnic groups within can lead to unintentional algorithmic bias. We chose not to focus on Eastern Asians for this experiment, as there can be a large origin overlap between both races that we may not have been able to

account for. We hoped to elucidate the merits of South Asians independently, and so we chose to compare them to a group that many algorithms are optimized for [5], and who have lesser potential of cultural or origin overlap with South Asians. This study compares perception and facial expressiveness between South Asians and Caucasians.

2 Related Work

There are a range of common differences we expect to observe between South Asians and Caucasians, including variance in skintone, anatomical facial structure, and facial expressiveness to a shared emotional state. After the discovery of "basic emotions" by Paul Ekman [4], which had a unique set of characteristics that distinguished them from other emotions, the idea of universal expressions that all people possessed spread like wildfire. There was strong evidence to show that a universality of facial expressions mapping to the same emotions existed [1,2]. It was argued that these basic emotions were "universally recognized and produced by people of all cultures, regardless of nationality, ethnicity, race, gender, age or religion" [3]. Ekman went on to create a Facial Action Coding System of different Action Units that each represented the degree of a certain facial muscle's movement [1,2]. His team then stressed the importance of specific action units that were used in the display of his "basic emotions", for example the use of Action Units 4, 5, 7 and 23 to be most commonly associated with the emotion of anger [6]. However, cultural differences were not heavily considered at the time, and Ekman developed an emotion recognition test by only using Caucasian actors [7]. This later received backlash, as more studies came out that looked into the differences between cross-cultural facial expressions [8,9], and cross-cultural recognition of emotional expression [10,11]. Furthermore, it has been addressed that gender and skintone algorithmic bias found itself into widely-available, facial analysis software [5]. With the release of open-source, general-purpose facial analysis libraries such as OpenFace, researchers can now test algorithmic bias across computational tools with greater ease [12,14,15]. Finally, we expect datasets titled "Asian" to likely be biased with a greater amount of Eastern Asian subject data. This notion comes from how survey data from 2016 showed how unlikely it was for various racial groups to consider South Asians as "Asian" or "Asian American" [13]. We hoped to explore the range of these cultural differences in both facial display and recognition.

3 Experimental Study

For this study, we sought to understand what differences, if any, exist between ways that a computational software may rate facial expressions, and how in-group and out-group members of two different ethnic backgrounds perceive and display facial expressions. We focused the study on two specific groups, South Asians and Caucasians, for reasons expressed in the introduction. To limit confounding variables, we also focused exclusively on anger. We chose to focus on a negative emotion because we believed the emotional state would be more likely to have pronounced muscular differences that could be detected both by people and software alike. We hypothesized that anger would be easier for our participants to pose in than sadness or disgust. For the computational tool, we used OpenFace, a state of the art, affect-recognition software.

Given these choices, we set out to answer a few specific questions:

1. Are there differences in the ways that South Asians and Caucasians display anger?
2. Are there trends across how South Asians or Caucasians perceive each other's level of anger?
3. Do South Asians and Caucasians perceive anger in the same way that they display anger?
4. Will OpenFace find higher levels of AU intensities in photos identified as very angry by humans?

To answer these questions, we conducted an experiment in three stages. (1) We collected two portrait photos of each South Asian and Caucasian participant posing with a completely neutral expression and with an angry expression. (2) We conducted a survey to determine how South Asians and Caucasians perceived the intensity of anger in these photos. (3) Lastly, we ran our photos through OpenFace and collected the resulting AU measures for later analyses.

3.1 Part I – Portrait Photos

We had read that longer focal lengths tended to make a subject appear smarter, more attractive and less approachable [16]. To avoid variation in these perceived affective properties, we ensured all photos were taken with a 55mm lens by two members of our research team with prior experience in photography. We wanted representation from: South Asian females raised in South Asia, South Asian males raised in South Asia, Caucasian females raised in the US, and Caucasian males raised in the

US. We ensured our participants fell into one of these four categories. No mixed-race participants were included, nor were South Asians raised outside of South Asia or Caucasians raised outside of the US. For reference, when we refer to South Asians or Caucasians during the rest of this paper, we are referring to South Asians raised in South Asia and Caucasians raised in the US.

Twenty-one (21) subjects of Caucasian or South Asian descent participated in the first stage of the experiment. Eleven (11) were male and ten (10) were female with a mean age of 23.7 years (SD=3.8). The average age of the female participants was 23.3 years (SD=4.9). For Caucasian females, the average was 21.2 years (SD=0.7), and for South Asian females, the average was 26.5 years (SD=6.9). Of the ten (10) female participants, four (4) were South Asian, and six (6) were Caucasian. The average age of the male participants was 24.1 years (SD=2.7). For Caucasian males, the average was 23.2 years (SD=1.8), and for South Asian males, the average was 24.8 years (SD=3.2). Of the eleven (11) male participants, six (6) were South Asian, and five (5) were Caucasian. All participants were Northwestern University students, ranging across undergraduate and graduate students. All photos were taken on Northwestern University's Evanston campus, where participants were found. A few participants were contacted by our team, to ensure a balanced representation of all four categories, while most others were randomly selected.

Participants were first asked for their consent to have their photo taken for a study on algorithmic bias. If they accepted, we asked them to fill out a form that asked for the following information: their initials, contact information, age, gender, race, race of parent 1, race of parent 2, where they grew up, and confirmation of consent to have photos taken and used in our study. We asked our subjects to pose with a completely neutral expression for their first photograph, as a baseline for what to expect in differentiating from an emotional expression. For the second photo, we asked participants to pose with an angry face. We allowed participants as much time as was necessary to compose themselves and, for the angry condition, to think of something that made them angry, and replicate how they would express anger in that situation.

While we took multiple photos, we chose the best quality, neutral and angry photos per participant. These photos were preprocessed before being inserted into the survey.

3.2 Part II – Surveys

For our survey participants, we also wanted representation from South Asian females raised in South Asia, South Asian males raised in South Asia, Caucasian females raised in the US, and Caucasian males raised in the US, to maintain consistency.

Forty-seven (47) subjects of Caucasian or South Asian descent participated in the second stage of this experiment. Twenty-five (25) were male and twenty-two (22) were female with a mean age of 23.9 years (SD=3.8). The average age of the female participants was 24.6 years (SD=4.7). For Caucasian females, the average was 22.5 years (SD=1.6), and for South Asian females, the average was 26.7 years (SD=5.8). Of the twenty-two (22) female participants, eleven (11) were South Asian, and eleven (11) were Caucasian. The average age of the male participants was 23.2 years (SD=2.7). For Caucasian males, the average was 23.6 years (SD=2.7), and for South Asian males, the average was 22.9 years (SD=2.7). Of the twenty-five (25) male participants, fourteen (14) were South Asian, and eleven (11) were Caucasian. Participants were selected from on and off campus, and from geographically diverse places. The surveys were sent to participants for them to take on their own, to ensure they had sufficient time and no pressure. All participants were contacted by our team.

Because our participants took the survey independently, we used a within-subjects design. Participants were randomly split into three groups, and three separate surveys with a unique set of photos (also distributed across the three surveys randomly; corrections were made to ensure that a neutral and an angry photo of the same person were not present in the same survey) were sent out to participants. This was done to ensure that the survey was a manageable size, to ensure that our participants did not lose attention during the survey. Creating one survey would have required a great deal of attention (21 photos * 1 question per photo = 21 questions), and we did not want our results to be skewed by distraction or boredom.

Figure 1: A sample question in one of our surveys.

Participants were allowed as much time as necessary. Participants first read a brief instructional paragraph that described their task but provided no information about the study. We then asked them to fill out a form that asked the following questions: age, gender, race, race of parent 1, race of parent 2, where they grew up, and confirmation of consent to have photos taken and used in our study. Then the experimental stimuli were presented in a random order, and the same question was asked regarding each photo: "Rate the level of Anger expressed by this actor:". The ratings for anger were done on a scale of 1 to 10, with 1 labeled as "No expression of Anger" and 10 labeled as "Very intense expression of Anger."

3.3 Part III – OpenFace

OpenFace is a free, open source face recognition software available on GitHub for developer use. OpenFace can analyze various factors, including facial landmark detection and tracking, head pose estimation, facial action unit recognition, and eye-gaze tracking. OpenFace allows for both real-time performance from a webcam and for uploading photos or videos.

In brief, OpenFace works by detecting a face using a pre-trained model. It then transforms the face into a format that is compatible with the neural network, using a transformation ability to estimate real-time posing. Lastly, it uses the deep neural network to embed the face on a 128-dimension unit hypersphere, which is a generic representation of a human face. The Euclidean distance created from this analysis makes it easier to cluster, identify similarities, and complete classification tasks than with other facial recognition techniques.

OpenFace used the following datasets for training AUs: Bosphorus, BP4D from FERA2015, DISFA, FERA2011, SEMAINE from FERA2015, UNBC, and CK+. During the training phase, 500,000 were passed through the neural net from CASIA-WebFace, which has 494,414 images of 10,575 individuals and FaceScrub, which has 106,863 images of 530 celebrities.

We conducted research on various facial recognition software, and eventually ended up choosing OpenFace because we were particularly interested in how we could compare FACS (Facial Action Coding System) results with human ratings of faces. FACS is a system that categorizes human facial movements by detecting individual muscular movements from a facial appearance. FACS allows a uniquely detailed view into Action Units (AU) that create an expression, and can detect the following AUs: 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, and 45. OpenFace provides both the presence of an AU and its intensity on a 0 to 5 point scale.

An important note is that OpenFace excels in detecting changes in facial movements, which makes it particularly suited for videos rather than photos. Since our photos are a snapshot and likely a mixture of multiple emotions, we may have introduced noise with the AUs, which we noticed in certain cases like with AU 6 (the Cheek Raiser, usually indicative of happiness).

4 Results

4.1 Q1) Are there differences in the ways that South Asians and Caucasians display anger?

In our analysis, we compared AU output from OpenFace and analyzed the differences between various categories to find trends: (1) angry vs. neutral faces overall, (2) angry vs. neutral faces in Caucasians, (3) angry vs. neutral faces in South Asians, (4) South Asian angry faces vs. Caucasian angry faces, and (5) South Asian neutral faces vs. Caucasian neutral faces. We used an independent t-test, a measure of inferential statistic used to indicate a significant difference between the means of two groups with a p-value of $p < 0.1$; all further analysis in this section used the p-value to compare the groups across these categories and determine the AUs that had significance for each category.

Our main findings were that all categories use AUs 4, 9, and 10 while showing anger, but that South Asians use AUs 9 and 10 both during anger and during the neutral condition. We also found a few AUs that were found in one cultural condition but not the other, such as AU 23, which was found more in Caucasian faces. Lastly, we found that less AU 5 was used to display anger, which may have implications on the perceptions of anger.

The following AUs will be discussed in this section: AU 4, AU 5, AU 6, AU 7, AU 9, AU 10, and AU 23. The AUs most commonly associated with anger are AU 4, 5, 7, and 23. AU 4 is the Brow Lowerer; AU 5 is the Upper Lid Raiser; AU 6 is the Cheek Raiser; AU 7 is the Lid Tightener; AU 9 is the Nose Wrinkler; AU 10 is the Upper Lip Raiser; and AU 23 is the Lip Tightener.

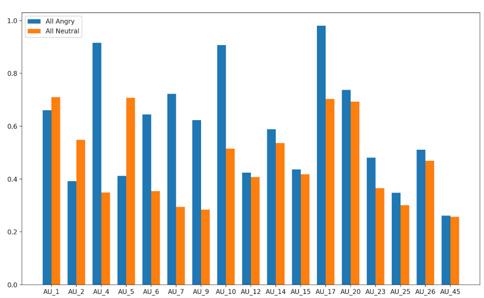

Figure 2: A graph indicating AU output from OpenFace, aggregating South Asian and Caucasian angry faces to identify the AUs most present in each category.

We found that in each of the five analyses conducted above, all categories use AUs 4, 9, and 10 while showing anger. Caucasians in particular show AU 4 at a more pronounced level than South Asians do while showing anger (as evidenced by the results in category (4), $p=0.095$) whereas AUs 9 and 10 were more pronounced in South Asian faces while they showed anger (category (4), $p=0.099$ and $p=0.077$, respectively). Interestingly, South Asian faces showed AUs 9 and 10 more than Caucasians in the neutral condition (category (5), $p=0.056$ and $p=0.043$, respectively), suggesting that AUs 9 and 10 are more pronounced on South Asian faces in comparison to Caucasian faces. This suggests a possible level of differences in South Asian and Caucasian faces and could have implications on the ways that each category displays and perceives emotions in the other category's faces, as discussed later.

AU 7 was found pronounced in the angry vs. neutral overall category (category (1), $p=0.076$) but not in any other comparison categories. AU 23, the lip tightener, was shown only in South Asian angry vs. Caucasian angry (category (4), $p=0.067$) with Caucasians showing this to a greater degree.

We also found AU 6 and AU 15 at fairly high levels across all variations of our analysis, but we consider this to be a confounding factor. It was clear during our data collection that our participants were surprised by our request to pose with a completely straight face or angrily, and that some felt awkward doing so; many of our participants laughed at this request before composing themselves and posing. We gave our participants as much time as they needed to regain composure and prepare their impression of anger, but since our participants laughed before posing angrily, we found pronounced levels of AU 6 (the Cheek Raiser, often associated with happiness) in all levels of our analysis. This makes sense, as much research that emotion does not appear in isolated events on the face, and that emotions instead occur in more of a blur towards each other.

Lastly, AU 5 was higher in the neutral condition as shown in the angry vs. neutral overall category than in any other category (category (1), $p=0.082$). This is interesting because AU 5 is an AU traditionally associated with anger, and we found that a lower level of AU 5 was correlated with angry faces. We were intrigued by the ways people chose to display anger. It was clear that anger was associated with some of the AUs we expected (4, 5, 7, and 23), because many of our participants lowered their brows, creating a slightly furrowed brow, tightened their eyelids, creating a squinting effect, and tightened the muscles around their lips, creating a pursed lip look (which were analyzed as AUs 4, 6, and 23). However, although AU 5 (the Upper Lid Raiser) is associated with anger, we found that people exemplified anger by squinting, which resulted in a decreased level of AU 5. We found this interesting because whether or not it actually shows anger, it provides an interesting insight into the ways people overall attempt to display anger.

4.2 Q2) Are there trends across how South Asians or Caucasians perceive each other's levels of anger?

With the intention of finding differences in how cultures rated one another's emotional states, we filtered through our survey data from 47 raters to find variance between South Asian and Caucasian raters on both neutral- and angry-photographed South Asians and Caucasians. Immediately, we noticed photographs of 2 individuals from the "neutral" category as being rated with greater "anger" scores than their photography from the "angry" category. Additionally, due to the subtlety of our participants' posed anger expressions when asked to "express how [they] look angry," many other neutral photos of individuals had similar, but still lower, scores to their angry counterparts. This unexpected finding showed that some people were rating neutral faces as angry faces, and we hypothesized that perhaps seeing a stranger with a completely emotionless face is an unfamiliar stimulus for the user, and that they may have perceived an emotionless face as anger.

We calculated the ratings for each neutral and angry photo of Caucasians to find that on average, South Asian raters scored both neutral and angry South Asian photos with higher "anger" scores than how Caucasians rated them. Similarly, Caucasian raters scored both neutral and angry Caucasian photos with higher "anger" scores than how South Asians rated them.

With the intent of finding within-group differences, we looked at the spread of scores (standard deviation) for each race. We expected ethnic groups to recognize emotional states of their own race with greater cohesion. This meant that we expected Caucasians to rate neutral and angry Caucasian photos with smaller standard deviation than how South Asians rate neutral and angry Caucasian photos, and vice versa. We did not find these results, and in fact, these values often reflected the opposite result of higher standard deviations.

Finally, we look at between-group differences. We expected Caucasians to rate neutral & angry Caucasian photos with more differentiable "anger" scores than how South Asians viewed their facial expressiveness, and vice versa. While this wasn't true when rating Caucasian photos, both groups rated South Asians very differently. We conducted an independent T-test to reveal significance between both groups rating all South Asian photos and angry South Asian photos (Caucasians and South Asians rating South Asians, p-value=.005; Caucasians and South Asians rating 'angry' South Asians, p-value=.007). Figure 3 shows a box plot (swarm plot) representation of the two groups' ratings on the aforementioned photos.

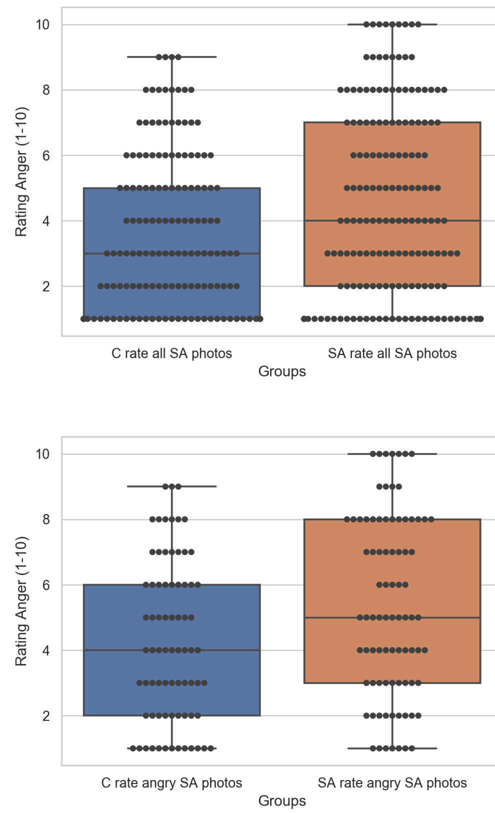

Figure 3: Box-plot representations of Caucasians and South Asians rating South Asian photos (p=.005 for neutral & angry photos (above); p=.007 for angry photos (below)).

4.3 Note on Q3 and Q4

Using our photos, survey results, and OpenFace output, we were able to detect differences in the ways that Caucasians and South Asians display and perceive faces in Q1 and Q2. These findings were based entirely in our dataset and were found with statistical significance. However, we were interested in more philosophical hypotheses, and have attempted to answer these in Q3 and Q4 as well. Our results are less statistically significant but provide an interesting ground for discussion regardless.

4.4 Q3) Do different cultures perceive anger in the same way that they display anger?

In our analysis, we looked at the photos that were rated most angrily (above 7 points on a 10-point scale) and least angrily (less than 3 points on a 10-point scale). Our hypothesis was that we would find higher degrees of the AUs used to display anger within a racial category (e.g., if Caucasians use AU 4 most to display anger, photos that were rated with a high anger score would also have a high level of AU 4). This relies on the key assumption that OpenFace is able to detect action units correctly and unbiased across cultures. In our analysis, we left in photos that were intended to be from the neutral condition, because we wanted to also understand what AUs were present in a neutral face to make participants perceive it as angry.

We were able to find similarities in the AUs used to display anger and those that were found in faces rated high on our anger rating scale. One of our main findings was that AU 4 was present in photos that both Caucasians and South Asians rated as angrier in all other categories. We also found an increase in AU 5 in photos of Caucasians that were rated less angrily by South Asians. Finally, we found AU 17 to be present in photos that both Caucasians and South Asians rated as angrier in all other groups, which is interesting because both Caucasians and South Asians showed high levels of AU 17 while displaying anger.

The following AUs will be discussed in this section: AU 4, AU 5, AU 12, AU 14, AU 17, and AU 26. As mentioned, the AUs most commonly associated with anger are AU 4, 5, 7, and 23. AU 4 is the Brow Lowerer; AU 5 is the Upper Lid Raiser; AU 12 is the Lip Corner Puller; AU 14 is the Dimpler; AU 17 is the Chin Raiser; and AU 25 is the Lips Part.

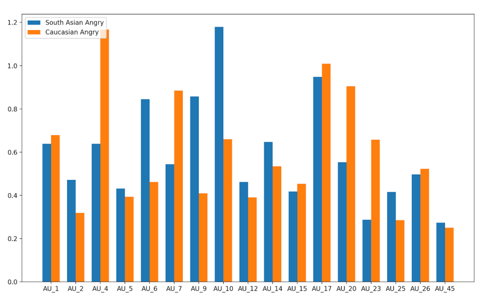

Figure 4: A graph indicating the AUs present in photos that were rated most angrily by participants in the survey task.

We found that AU 4 was present to a statistically significant degree in photos that Caucasians rated as angrier in all other categories (p=0.028) and photos that South Asians rated as angrier of all other categories (p=0.096). This makes sense and is an interesting observation given the prevalence of AU 4 in displaying anger across all racial groups and categories. This could have implications regarding the similarity between factors used to display and perceive anger and could suggest that AU 4 is an important differentiating factor both in display and perception of anger across Caucasian and South Asian categories.

Our results also showed that an increase of AU 5 was present to a statistically significant degree in photos that South Asians rated less angrily of Caucasians (p=0.007). This could mean that when Caucasian individuals use AU 5, which creates a squinting effect, South Asians view it as anger more than other categories do. This is also consistent with our results from H1, as an increase in AU 5 was found in the neutral conditions, and a decrease in AU 5, or increased amount of squinting, was present in the anger conditions; however, in H1, these results were found across the board, whereas in this condition, they were only found between Caucasians and South Asians.

We also found that AU 17 was present to a statistically significant degree in photos that both Caucasians and South Asians rated as angrier in all other groups (p=0.052). This is a very interesting result, because AU 17 was also the AU that both South Asians and Caucasians had in common to the highest degree while displaying anger. This could indicate that perhaps AU 17, the Chin Raiser, is thought of as a common way to show and identify anger in Caucasians and South Asians. Note that AU 17 is not part of the previously mentioned AUs associated with anger: AUs 4, 5, 7, 23, but perhaps this has implications on the ways that anger is perceived while being posed, rather than actually shown when elicited.

On a similar note, our results also showed that an increase of AU 12 was present to a statistically significant degree in photos that Caucasians rated less angrily of South Asians (p=0.09). Again, we attribute the presence of AU 12, traditionally associated with happiness, to the fact that our participants were not trained actors, and likely had some indication of happiness on their faces after having laughed moments before the photograph was taken. However, the association of AU 12 with less anger is logical, since AU 12 is associated with happiness.

4.5 Q4) Will OpenFace find higher levels of AU intensities in photos identified as very angry by humans?

A final question we sought to explore was how OpenFace's readings of emotion compared to that of a human's readings of emotions. We specifically looked to see if OpenFace's readings of anger intensity maps to the intensity of anger rated by human subjects. To perform this analysis, we looked to see if OpenFace was able to notice significant differences in photos rated as "very angry" and "not angry" by humans.

The first step to this analysis was to separate photos as "very angry" or "not angry". This required us to both choose an "accurate" rating for the photo, and also to determine a threshold for choosing if that rating is either "very angry", "not angry", or neither. To choose the most "accurate" rating, we used ratings of South Asian photos by other South Asians, and vice-versa for Caucasians. This selection was based on a key assumption that intra-cultural ratings are more accurate than intercultural ratings. Making this assumption allows us to use these ratings as the "gold standard", or ground truth, to compare OpenFace results. Finally, our criteria for a photo being "very angry" was that it was rated as having an average anger intensity above 7.0 and below 3.0 for "not angry". This was chosen to remove all photos viewed as "neutral."

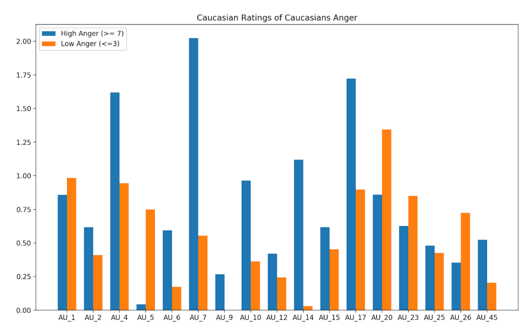

Figure 5: A graph indicating the AUs present in photos of Caucasians separated by ratings of Caucasians.

We begin with an exploration of OpenFaces ratings on the Caucasian group (Figure 5). In the figure, you are able to see the average action unit intensity for "very angry" photos in blue and "not angry" photos in orange. As noted in the related works, traditional mappings from action units to the anger emotion are through changes in Action Units 4, 5, 7, and 23. While there are no statistical differences for differences between groups on these units, we can still see general trends that show OpenFace is recognizing some change in intensity of anger. AU 4, 5 and 7 change in the "High Anger" case by almost a point, while 23 has a negligible change. This leads us to lean toward the fact that while OpenFace isn't detecting changes in intensity at a statistically significant correctness compared to Caucasian ratings, it is following the correct general trends.

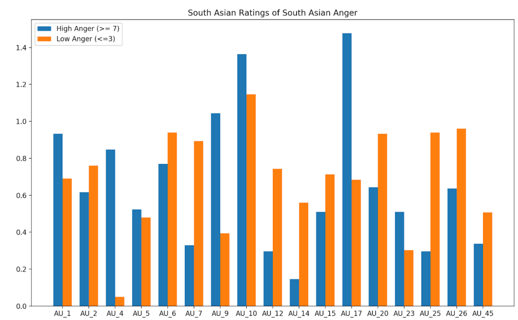

Figure 6: A graph indicating the AUs present in photos of South Asians separated by ratings of South Asians.

This cannot be said, however, for photos of South Asians rated as angry by other South Asians. When we look at Figure 6, while also not statistically significant, we start to see troubling trends in the Action Unit Mappings. While AU 4 does increase in the photos rated as high anger by South Asians, does not go above an average of 1.0. AU 5 does not change much at all, and the small change it does exhibit is in the opposite direction that we want (AU5 should decrease for angry photos). AU 7 has a significant change, but also trends in the wrong direction. AU 23 trends in the correct direction but does not show a large change in means.

We start to see from each of these graphs that OpenFace traditional emotional readings do not match the readings of intercultural ratings. The readings are much better for Caucasians than for South Asians, but there is no statistical guarantee that even OpenFace readings of Caucasians are correctly mapped. This leads to even further evidence that different cultures use different AUs in the display of emotion, as explained in H1, creating a need for a potentially more unique and nuanced group of anger related AU's. For example, maybe mapping different AUs to different races to more accurately determine an emotion based on the ethnicity of a face could be helpful.

5 Discussion

A key point of discussion for this paper's findings and for future findings will have to include the debate on what is a "correct" or "accurate" reading of emotion. As our study has shown, there are numerous discrepancies between how tools, people, and people of different cultures understand emotion. Making claims beyond these differences, as we have explored in section Q3 and Q4, however, requires strong assumptions of "who is right", or an understanding of some ground truth to compare data to. For our analysis, we made key assumptions about how to judge the findings of OpenFace, such as the idea that "intra-cultural ratings are more accurate than inter-cultural ratings' or that "OpenFace is able to correctly able to detect action unit changes for different cultures." However, these assumptions could be challenged in a myriad of different ways, and should be, to get a greater understanding of the effects of cultural differences on technology.

Regardless of the assumptions made, findings of cultural differences have great ethical implications. Using computational tools for facial recognition in e-commerce, policing, or even simply unlocking a phone without considering their effectiveness across cultures has broad implications that can range from criminal justice to designing for equality. Thus, as these tools gain more acceptance and usage, it is critical that they are able to provide nuanced and unique readings for different cultures to provide value and fair treatment of all people. In this introductory study, we have shown differences between two cultures with one software, but it is likely that these differences will become pronounced across more cultures and with more tools.

Moreover, even beyond Action Units and expression, further work will also need to be done to understand different inherent differences between cultures. Action Units are just one piece of the puzzle of affect recognition. Differences in actual face structure, gaze patterns, and pose are all different factors that could affect the perceptions and displays of emotion across cultures. Understanding a full picture of these differences is a large task, but this type of work is needed to continue to shape our tools in a more holistic way.

6 Limitations

The findings from this study indicate that there are differences in the ways that South Asian and Caucasian people display emotion on their faces and perceive emotion from people outside their racial category.

This project was an initial exploration of in-group affect recognition and computational software, and due to its scope, we acknowledge many assumptions and limitations that make excellent ground for future work. One major limitation was the pool of participants we were able to recruit from. Firstly, given that many of our participants were found on the Northwestern University community's campus and that many participants were contacted by our research team, there is a likelihood that there is a degree of familiarity between participants, particularly those of South Asian descent. This could skew the results, because a person familiar with another person's face likely would be able to better gauge the emotion displayed on their face. We attempted to mitigate this risk by recruiting some participants off-campus for the survey task. Similarly, nearly all participants were in the same age group (approximately 21-29 years old), since most participants were students. The results may have been different if participants had to rate a face in a completely different age group. Due to the scope of our study and the limited participants we had access to, we felt that these were necessary tradeoffs, but we would like to work on recruiting a more diverse participant pool in the future.

We would also like to note that our photos and the survey were taken in the wild as opposed to a lab environment. Given our participant pool and no financial incentive available to offer them, we thought this method would allow for the largest number of participants possible and that this would be a more comfortable environment for our participants to pose with an emotion. However, this may have created varying lighting and photo quality conditions, which could have influenced the OpenFace results. We also had our participants pose with an emotion rather than eliciting an emotion in order to analyze the different ways different racial groups display emotions, but we recognize that perhaps participants did not pose in a manner that accurately reflects an angry expression, and that this could have affected the recognition task asked of other participants. As mentioned earlier, we found AU 6 and AU 15 at fairly high levels across all variations of our analysis, but we consider this to be a confounding factor, and a limitation in our dataset.

Lastly, we noticed some limitations with the scoring produced in the survey task. Some of our participants rated most photos on extremes, either extremely angry or extremely lacking anger. This could have skewed our data since much of our analysis is on averages and medians, which can be skewed by a large amount of data on the extremes. However, this also reveals the possibility that participants perhaps do not see emotion on a scale and find the task of rating a level of emotion using numbers difficult and prefer a more binary scale.

We view these limitations as an opportunity for a more thorough future study and encourage other teams to continue exploring in-group and computational affect recognition.

7 Conclusion

In this study we conducted a thorough analysis of the differences and similarities in perception and display of anger between South Asians and Caucasians. Using Open Face's Action Unit Landmarking tool, we were able to display that South Asians and Caucasians have different points of emphasis on their face that express anger. Using our survey data, we were able to interpret that South Asian's anger is perceived significantly differently by Caucasians and South Asians. We were also able to begin an analysis on the ability of OpenFace to detect changes in intensity of anger across cultures. We believe our findings continue an ongoing discussion and movement towards building our computational tools in a more inclusive and culturally diverse way. Our data set of photos and survey responses is available for open source use, and we hope it is used to continue research in this direction.

REFERENCES

[1] Ekman, P., & Friesen, W. V. (1978a). Facial Action Coding System (FACS): A technique for the measurement of facial action. Palo Alto, CA: Consulting Psychologists Press.

[2] Ekman, P., & Friesen, W. V. (1978b). Facial Action Coding System: Part two. Palo Alto, CA: Consulting Psychologists Press.

[3] Matsumoto, David. The Origin of Universal Human Emotions . pdfs.semanticscholar.org/91b9/d6c7da6986d0b1a3aa066045cd7b104d2513.pdf.

[4] Ekman, Paul. "Basic Emotions." Handbook of Cognition and Emotion, 2005, pp. 45–60., doi:10.1002/0470013494.ch3.

[5] Hardesty, Larry, and MIT News Office. "Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems." MIT News, 11 Feb. 2018, news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212.

[6] "Facial Action Coding System (FACS) - A Visual Guidebook." IMotions, 18 Feb. 2019, imotions.com/blog/facial-action-coding-system/.

[7] Ekman, Paul. "Pictures of Facial Affect (POFA)." Paul Ekman Group, 1993, www.paulekman.com/product/pictures-of-facial-affect-pofa/.

[8] Ekman, P. (1972). Universals and cultural differences in facial expressions of emotions . In J. Cole (Ed.), Nebraska Symposium on Motivation, 1971 (pp. 207–283). Lincoln: University of Nebraska Press.

[9] Jack, Rachael E., et al. "Cultural Confusions Show That Facial Expressions Are Not Universal." Current Biology, vol. 19, no. 18, 2009, pp. 1543–1548., doi:10.1016/j.cub.2009.07.051.

[10] Boucher, J. D., & Carlson, G. E. (1980). Recognition of facial expressions in three cultures. Journal of Cross-Cultural Psychology , 11 , 263–280.

[11] Schimmack, U. (1996). Cultural influences on the recognition of emotion by facial expressions: Individualistic or Caucasian cultures? Journal of Cross-Cultural Psychology , 27 , 37–50.

[12] Baltrusaitis, Tadas, et al. "OpenFace: An Open Source Facial Behavior Analysis Toolkit." 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 2016, doi:10.1109/wacv.2016.7477553.

[13] Wang, Alton, and Karthick Ramakrishnan. "Who We Count as Asian American: #BrownAsians." Data Bits, 20 Nov. 2017, aapidata.com/blog/who-we-count-as-asian-american-brownasians/.

[14] "OpenFace- Facial Landmark Detection and Facial Expression Analysis Software." Flintbox, www.flintbox.com/public/project/50632/.

[15] Kim, Stephanie. "Understanding Facial Recognition Through OpenFace." Algorithmia Blog, 5 Dec. 2018, blog.algorithmia.com/understanding-facial-recognition-openface/.

[16] Banks, Martin S et al. "Camera Focal Length and the Perception of Pictures." Ecological psychology : a publication of the International Society for Ecological Psychology vol. 26,1-2 (2014): 30-46. doi:10.1080/10407413.2014.877284