BEX3012 Project Report Detecting Facial Expressions in Professional Tennis Matches

Stephanie Kobakian

12 May 2017

Contents

Tennis is often considered a mental game due to the pressures placed on the two or four individual during match play. This report explores the application of emotion recognition technology to elite tennis match broadcasts. This extends on research of the facial detection performance of these APIs. By understanding the emotions expressed on a face we have the opportunity to consider a player's mental state as a match progresses. This would help players to improve their performance if correlations can be found between match performance and their emotional state. By analysing a players mental state from observations rather than self reporting we may be able to gain information that a player will not recall after a match. It is predicted that the software will be able to derive the emotional expression shown by a player. It is expected that emotional recognition may be impaired by certain features of the face captured.

1 Materials

1.1 Images

The image set used to derive emotion information contained 6406 still shot images. These were 800x450px video frames, taken every 3 seconds from 5 minute video segments. These segments were sampled from 105 Australian Open 2016 match broadcast videos. The broadcast videos are the televised match play shown on Channel Seven. They contain shots from various cameras, these angles vary depending on the court the match was played on.

Figure 1: This image of John Millman was taken from the broadcast video of a match played against DSchwartzmann in the first Round of the 2016 Australian Open. This image is representative of the most commmon group of faces found. It captures a player's face close up on the court, taken at the player's shoulder height, with the logo wall in the back ground.

1.2 Software APIs

Three emotion recognition APIs were considered. They were chosen for their accessibility and online reviews of their performance. Each API is a well recognised Facial and Emotion recognition solution currently available online.

Table 1: This table details the capabilities that were considered in contrasting the accessibility and outputs of APIs that recognise emotions in images of faces.

Attribute	Google	Microsoft	Skybiometry
Call Limits	10 per second	20 per minute	100 per hour
Emotion Output	Categorical Likelihood	Numeric Proportions, (0.0-1.0)	Confidence Value, (0-100)
Number of Emotions	4	8	7
Cost and Access	Account and Payment	Account	Account and Payment
API Access	REST	REST	REST

The three APIs reviewed in this report have been summarised in Table 1 above. A noticeable difference between the three is the amount of times the API can be called within a given time frame. Skybiometry had the largest imposition on Call Limits as it only allowed 100 API calls to b e processed per hour. Microsoft also had a limit imposed, but this allowed for much more to be processed with the possibility of 1200 images to be processed within one hour, after accounting for the wait time between each group of 20. Google Vision's API call limit had a minimal effect for our purposes.

There were a different set of emotions in the results provided by each API. These differences made it necessary to process the outputs to provide the most comparable results for analysis. Microsoft provided the most emotional categories, 'contempt' being the addition that was not provided by Skybiometry or Google. Google provided the least, only providing four emotions, joy, sorrow, surprise and anger.

The observations of these emotions from the APIs were also different. Google provides likelihoods of an emotion occurring on a particular face in categorical options ranging from Very Unlikely to Very Likely. Microsoft provide Proportions, ranging between 0 and 1, the sum of the emotion values for each face summed to approximately one. Whereas Skybiometry results in a confidence value of the emotion occurring in the specified face in values from 0 to 100, these values were not mutually exclusive and resulted in a different sum across each face.

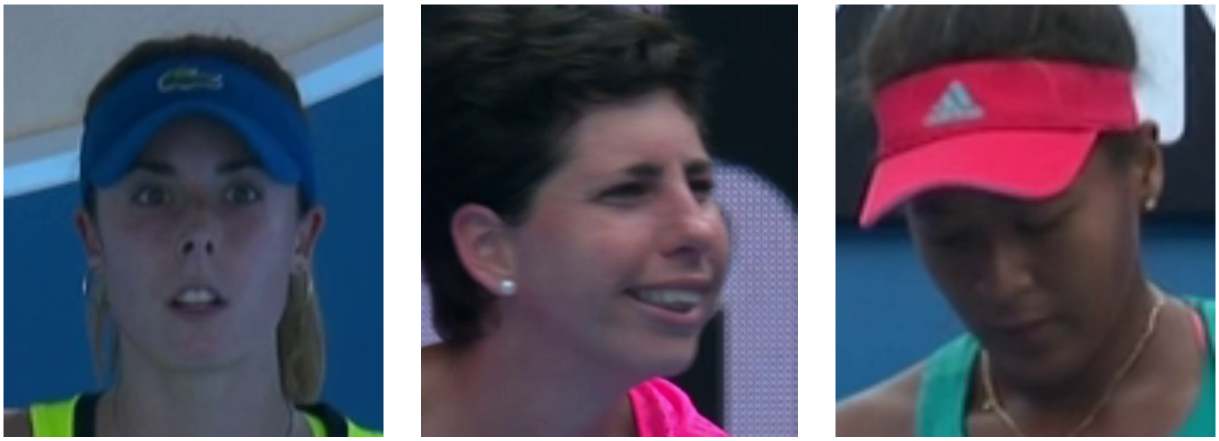
Cost and Access refers to the steps that must be undertaken for a new user to begin Emotion Recognition using each API. All options require an account be created with a current email address. Google and Skybiometry require payments to be made. The Google Vision API provides Facial and Emotion Recognition for $1.50 for Units 1001 - 5,000,000 per month[1]. Skybiometry have three API subscription options that vary in the amount of API calls that can be made per month, day and hour[2].

[1] https://cloud.Google.com/vision/docs/pricing
[2] https://skybiometry.com/pricing/

2 Procedure

2.1 Process Images

To create a set of individual faces for the APIs to consider we derived small images that were subsets of the 450x800px images. The crop was performed to focus on the area marked as a face by Google[3]. These 1319 new images, of varying sizes, were hosted on Google Drive as individual images.



Software	Emotion
Microsoft	surprise
Skybiometry	surprise
Google	neutral

Software	Emotion
Microsoft	happiness
Skybiometry	fear
Google	happiness

Software	Emotion
Microsoft	NA
Skybiometry	NA
Google	neutral

Figure 2: This figure provides three faces in the image set, after extracting faces from the full broadcast video stills. Below the faces are the APIs predominant emotion category recognised in the face.

2.2 Process API results

Given the different forms of the outputs produced by each API, we felt a harmonization process was required prior to undertaking a comparative analysis. The goal of the harmonization was to produce Emotion values for each face that would be comparable across the three APIs.

Microsoft provided a POST request object in a JSON format. This was easily manipulated to create a data set of the eight numeric emotion values provided for each face. In the event that emotion values were not returned for a certain face, a row of NAs is associated with the unique face ID in the data set. This data set contained a column for the following emotions: anger, contempt, disgust, fear, happiness, neutral, sadness, surprise.

[3]based on Google's bounding box coordinates

Table 2: The Microsoft API provided eight numerical values, one for each emotion.

anger	contempt	disgust	fear	happiness	neutral	sadness	surprise
0.000	0.000	0.000	0.003	0.000	0.136	0.000	0.861

Table 2 details the emotion category levels for the first face in Figure 2. The highest emotion level is for surprise, and the lowest value of 2.78e-05 is associated with contempt.

Skybiometry also provided a POST request object in a JSON format. However, unlike Microsoft, the results included a True or False Value for each of the seven emotions and a 'confidence value as a percentage from 0 to 100'[4] representing the confidence of the emotion being present on a player's face. Table 3 contains the confidence values for the first face in Figure 2. If a confidence value was 50 or above, the emotion 'value' would be labelled True. This would have occurred for surprise and disgust.

Table 3: The Skybiometry API provided seven numerical Confidence values, one for each emotion.

neutral	anger	disgust	fear	happiness	sadness	surprise
0	0	52	3	0	0	96

The process undertaken to receive the results Google produced was the same as the previous APIs. However rather than numeric values, it returned categorical likelihoods for the four emotions it considered.

Table 4: The Google Vision API provided four likelihood possibilities, one for each emotion.

happiness	sadness	anger	surprise
VERY_UNLIKELY	VERY_UNLIKELY	VERY_UNLIKELY	UNLIKELY

The same face that had produced high levels of surprise for Microsoft and Skybiometry produced the results in Table 4. The most positive result was the 'Unlikely' categorisation for surprise, this proposes that surprise was more prominent than other emotions but less significant than the other APIs intimated.

[4] https://skybiometry.com/documentation/#document-21

2.2.1 API Result Transformations

Noticing that the sums of the individual emotions for each face in the Microsoft results was approximately one, provided the basis for transformation for the other APIs. Dividing the individual Skybiometry emotion confidences by the sum of the confidences would give the proportional likelihood of each emotion being the dominant emotion in a face.

This calculation was simple to perform and resulted in numeric values comparable to Microsoft's values.

Skybiometry Transformed:

Table 5: The Skybiometry API provided 7 numerical Confidence values, one for each emotion.

neutral	anger	disgust	fear	happiness	sadness	surprise
0	0	0.344	0.02	0	0	0.636

Google's results then required a more complex transformation for the output to also be comparable.

As there were five possible categories assigned to each emotion the numeric values were chosen to sit in the middle of five 20 point ranges. This allowed a numerical representation of each possible likelihood value and would allow for numeric comparisons to be made. The values were assigned according to the table below:

Table 6:

Likelihood	Value
VERY_UNLIKELY	10
UNLIKELY	30
POSSIBLE	50
LIKELY	70
VERY_LIKELY	90

As both Microsoft and Skybiometry included a neutral category we incorporated a neutral category for the Google results.

Our approach considered that when a face was "very unlikely" to be any of the four emotion categories it could be deemed "neutral". However when one emotion was stronger than others it could still be possible that the face only had a small lean toward this emotion and it may not be a dominant expression . Therefore the Confidence value for neutral was calculated according to the formula below:

$$neutralConf_i = 100 - max(happinessConf_i, sadnessConf_i, angerConf_i, surpriseConf_i)$$

The process to derive the numeric values for Google followed the steps undertaken for Skybiometry values. Where the individual emotion confidences were divided by the sum of the emotion confidence values for each face. This transformation results in the values seen in Table 7, the faces can now be analysed numerical and compared to the other APIs numerical results.

Table 7: The Google Vision API provided five proportions, one for each emotion and a neutral category.

happiness	sadness	anger	surprise	neutral
0.077	0.077	0.077	0.231	0.538

2.2.2 Predominant Emotion Categorisations

A manual addition to our data set was a Dominant Emotion variable. This will allow comparisons of the emotion results for certain faces across the three APIs beyond comparing the individual emotion results for all emotions. For Google, Skybiometry and Microsoft, the emotion proportion levels for each individual face were considered, the maximum proportion value of each APIs emotion set was deemed the dominant emotion for each face. Recalling that there is a difference in the amount of emotion categories considered by each software.

For each individual face this resulted in three dominant emotion classifications, one per API. The dominant emotion would be one of seven emotions for Google, one of eight for Microsoft, and one of five for Google.

3 Results

Each API returned a different amount of total responses, where 'responses' are not NA results for emotions on a particular face. Microsoft produced emotion results for 732 faces. This was more than Skybiometry, as it produced only 634 results for emotions on faces. Google provided the most, with 1242 emotion results[5].

Table 8: Emotion results were found by all three APIs for 364 common faces, none of the APIs produced results for 10 player faces. Google was able to produce 380 results that Skybiometry and Microsoft did not.

		Skybiometry	FALSE	TRUE
Google	Microsoft			
FALSE	FALSE		10	1
	TRUE		4	18
TRUE	FALSE		380	53
	TRUE		113	364

3.0.1 Player Faces

Table 9: A logisitic regression was undertaken to show how certain image attributes contribute to the likelihood of the face being a player.

	exp(Estimates)	exp(2.5%)	exp(97.5%)
bgCrowd	0.001	0.000	0.008
bgLogo wall	0.049	0.003	0.267
shotanglePlayer Shoulder Height	4.453	2.210	9.051
shotangleUpward Angle	0.832	0.376	1.852
obscuredYes	0.634	0.428	0.939
headangleOther	2.278	1.343	3.871
headangleProfile	4.237	2.248	8.098
glassesYes	0.060	0.032	0.109

The most significant predictors of the face captured being that of a player are outlined in **??**. It can be seen that when the image was taken at a player's shoulder height then it is, on average, 4.45 times more likely to be a Player than a fan or staff member on court. When the head angle is Other, it is between 1.34 and 3.87 times more likely to be a Player than faces that have been captured front on, at the 5% level of confidence.

In the following sections we will only consider the faces of players, as they would be the intended targets for our applications of this research. Microsoft produced emotion results for 499 player's faces. Skybiometry, produced 436 and Google provided 910 emotion results for faces.

[5] As the initial set was faces identified by Google there is the possibility of bias toward being able to identify emotions in faces it was able to find originally.

3.0.2 Emotion Proportion Distributions

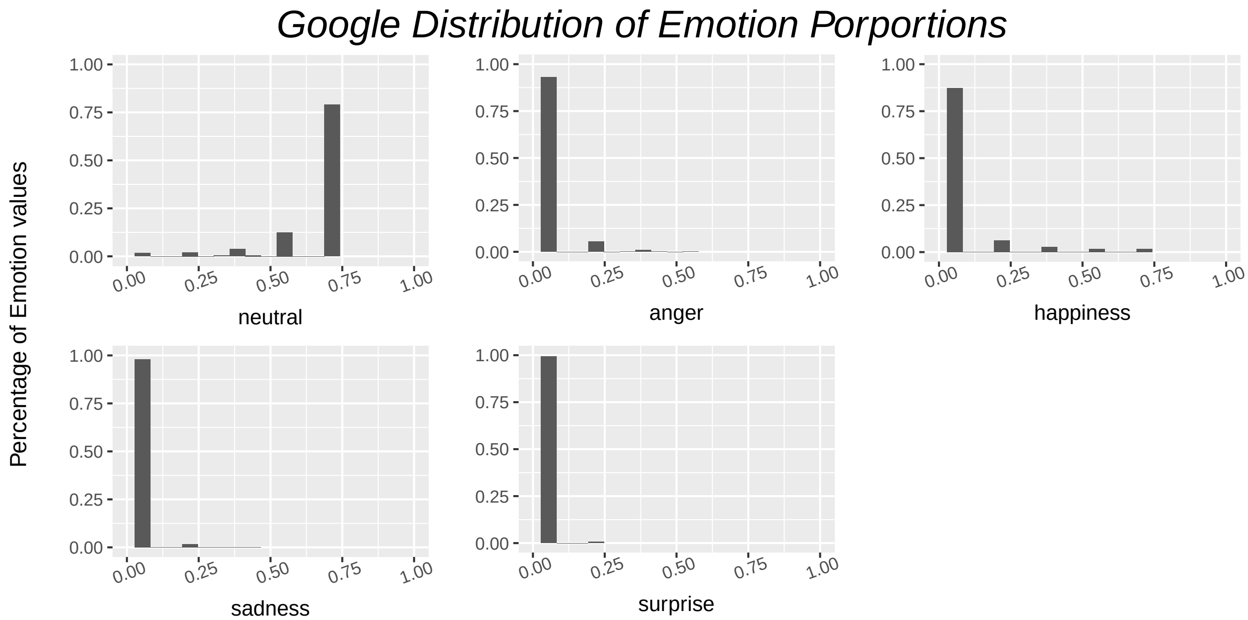


Figure 3: It can be seen that there are significant proportions of faces where the non neutral emotions occur close to zero. This is balanced by the large percentage, over 75%, where the neutral value of just under .75 occurred. The faces used showed higher levels of happiness than any other non neutral emotion as the percentage occuring at zero was less than the percentages of the other emotions.

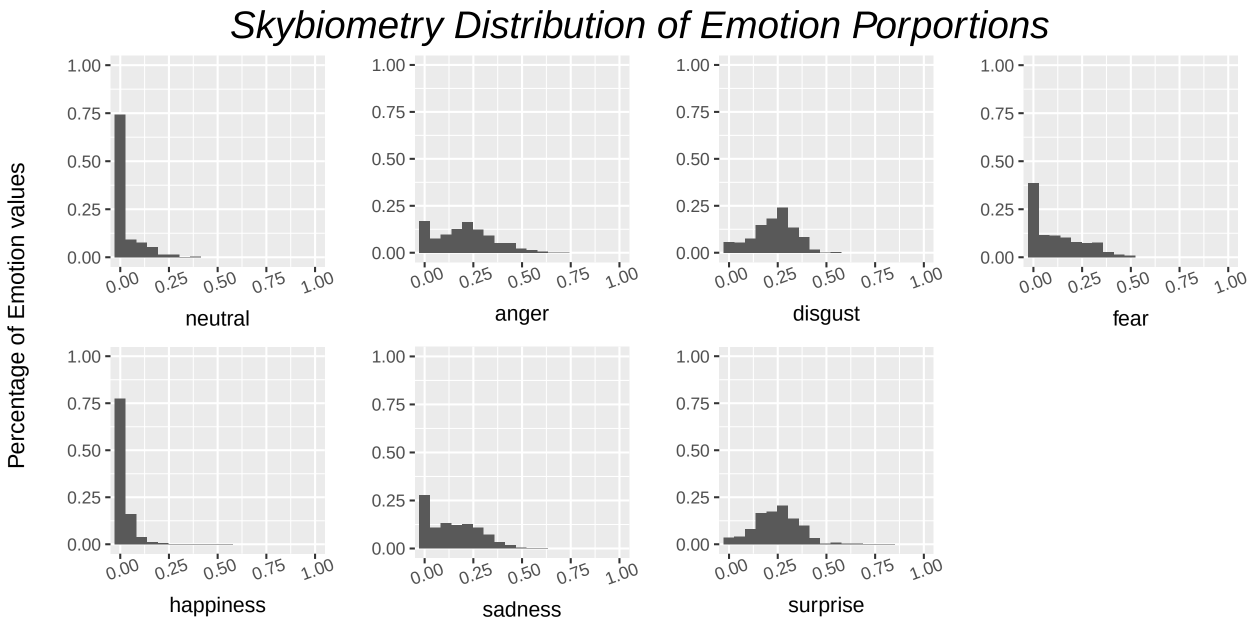


Figure 4: The distributions show that over 70% of faces had levels of happiness and neutrality at, or close to, 0. The distributions show that there are very little faces with specific emotion levels above 50%. Instead there are low levels of all emotions found on many faces.

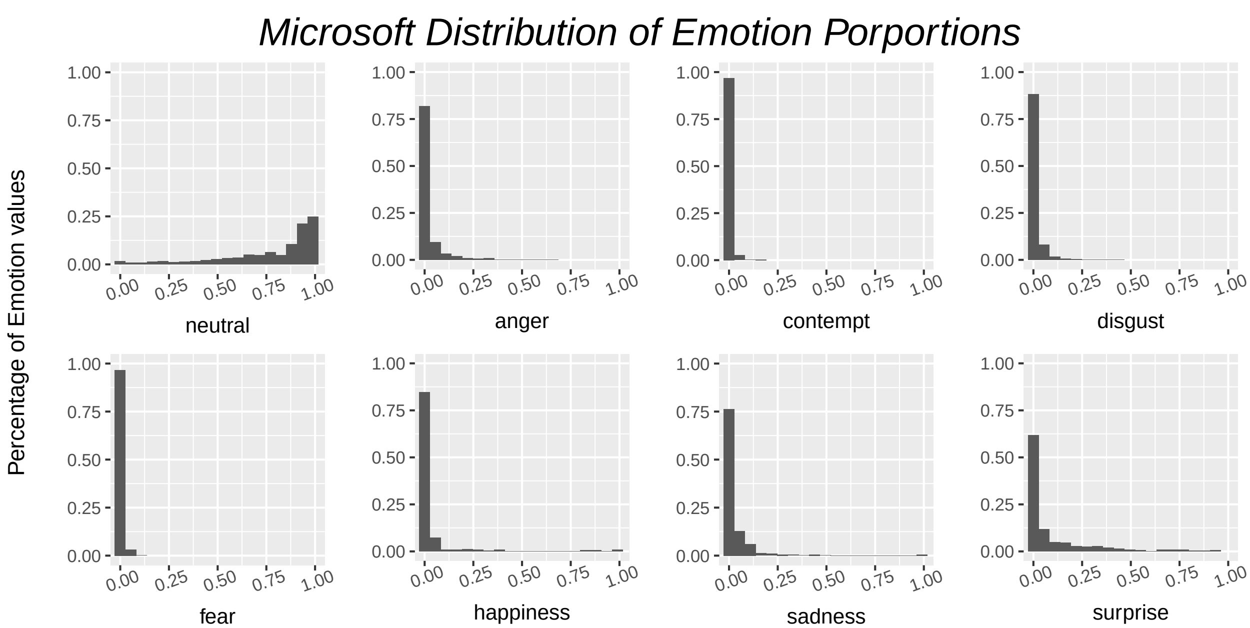


Figure 5: There were many faces that returned high neutral values. This is matched by the high percentages of faces where the proportion levels were at, or close to, zero. While the happiness emotion was most varied, with values between 0 and 1, over 40% of faces returned surprise levels above zero.

3.0.3 Predominant Emotions

Alize Cornet's predominant emotion in the first example image in Figure 2 has been considered by all three APIs and their classifications can be seen in the table below her image. Table 2, and Table 3, show the numeric proportions of all emotions for this image and they show that surprise is the predominant expression according Microsoft and Skybiometry. This differs from Google's classification as the Unlikely classification in Table 4 resulted in a high neutral value being incorporated into the emotion category proportions considered when allocating the predominant emotion to the face.

If we consider the predominant emotions of the faces considered by the Google API, given a reasonable number of faces, the Predominant Emotions would be expected to enforce the results shown previously in Figure 3.

As had the least amount of categories we have split the results according to the possible classifications of predominant emotions by Google.

The following tables have the predominant emotion classifications for Microsoft in the rows, and the columns denoted classifications by Skybiometry.

Table 10: Where Google provided the Emotional Result of Neutral; Microsoft associated high neutral values with many of these faces, however Skybiometry only classifed 3 of these faces as neutral.

	neutral	anger	disgust	fear	sadness	surprise	NA
neutral	3	82	62	46	37	72	97
anger	0	1	0	1	0	1	1
happiness	0	2	0	0	2	0	1
sadness	0	1	1	0	1	0	0
surprise	0	1	3	3	2	12	7
NA	0	14	8	3	6	20	348

This result supports the analysis of the numeric proportions found by Skybiometry in Figure 4. The three faces categorised as neutral must have had unusually high levels of neutral compared to the proportions of other emotions. This low number of occurrences is likely due to almost 75% of faces having a neutral level at 0.

Recalling that a neutral proportion was implied when Google did not allocate high likelihoods of emotions being present. The bias Google showed to allocating a neutral predominant emotion resulted from the high amount of low proportions of emotions, shown visually in 3. Over 75% of faces had a neutral level of 0.75. This is an extremely high amount of faces with little emotion recognised. This is why Table 10 above showing NA values for many faces when Google considered them to be predominantly neutral may imply that it is inclined to return results, even if these results are not informative.

Table 11: Where Google provided the Emotional Result of Happiness. The most common Skybiometry response was

	anger	sadness	surprise	NA
neutral	3	0	0	7
happiness	0	1	1	9
NA	2	0	0	34

It can be seen that when Google classified faces as Happiness Table 11 shows less faces to be considered by Microsoft and Skybiometry than Table 10 above when Google classified faces as Neutral. Many of these faces were not emotionally classified by Microsoft and Skybiometry resulting in 34 NAs. There were no faces considered to be displaying Happiness by all three APIs, however Microsoft did consider 9 faces to be happiness that Skybiometry did not return emotion results for.

These results were expected given that in Figure 3 the Happiness histogram showed almost 90% of faces with the lowest possible level of happiness. Leaving just over 10% to be showing even a small level of the likelihood of happiness appearing on the face.

Table 12: Where Google provided the Emotional Result of Anger

	anger	sadness	NA
neutral	5	2	5
NA	1	0	0

When Google provided a result of Anger, the only predominant emotion supplied by Microsoft was Neutral. It is a very a small group of faces, 13 of the set of 1319, and 5 of these were also considered to be anger by Skybiometry.

The difference between the amount of neutral predominant categorisations by Microsoft and Google in comparison to Skybiometry lead to further investigations into neutral compared to non neutral predominant categorisations.

As neutral was a prominent category for Microsoft and Google, we considered two groups in the following analyses, one where neutral was the predominant emotion and the other group was created by combining the remaining emotions into a non-neutral category.

Table 13: This shows the agreement between non-neutral and neutral faces recognised by Goole, Microsoft and, Skybiometry. Only three faces were categorised as neutral. Of these, Google was much more likely to categorise as neutral.

		Skybiometry	neutral	non-neutral	NA
Google	Microsoft				
neutral	neutral		3	299	97
	non-neutral		0	31	9
	NA		0	51	348
non-neutral	neutral		0	21	3
	non-neutral		0	10	4
	NA		0	2	32
NA	neutral		0	14	4
	non-neutral		0	4	0
	NA		0	1	10

Table **??** above shows that Skybiometry is less likely to categorise a face as neutral than Microsoft or Google. It can also be seen that 348 faces were considered neutral by Google but no emotional information was gained from Microsoft or Skybiometry. A large cross section category is where Google and Microsoft categorised a face as neutral and Skybiometry considered it to be non neutral, this happened for 299 faces. There is only one face that was found by Skybiometry to be non-neutral that Google and Microsoft did not return emotional information for.

Focusing only on the neutral proportion levels for every face leads us to examine the following. Figure 6 displays a Scatter Plot Matrix that captures the pairwise variability in the Neutral proportion levels across APIs.

We consider that there were attributes of the faces captured that may have influenced the likelihood of a neutral result in the cases of Google and Skybiometry. To understand this we consider the associations between the different attribute values and the neutral results.

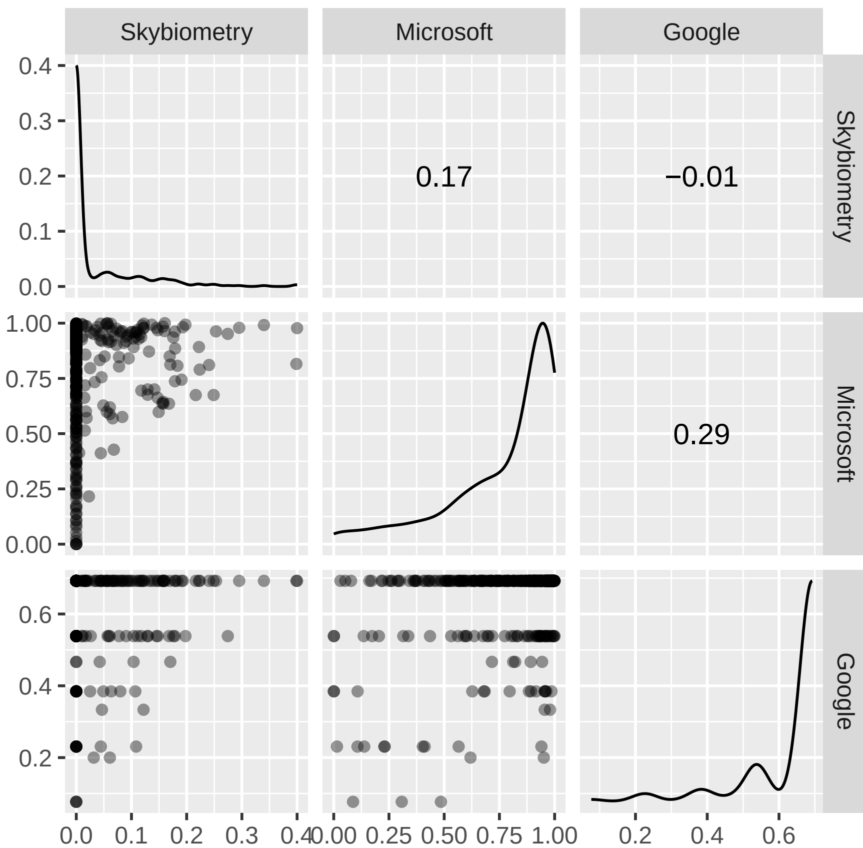


Figure 6: The Scatter Plot Matrix shows the variablitiy in the Neutral proportion levels. The -0.01 correlation value for Google and Skybiometry show that they do not behave in a similar way, or inversely. They are uncorrelated. Some high values of neutrality for Google are matched to some high levels found by Skybiometry. However, in the lower ranges of both there is a lot of variability in proportion values.

Table 14:

	exp(Estimates)	exp(2.5%)	exp(97.5%)
bgCrowd	0.077	0.004	0.436
bgLogo wall	0.177	0.010	0.875
obscuredYes	1.953	0.975	4.275
glassesYes	0.099	0.034	0.279

When a player's face is obscured it is, on average, 1.95 times more likely to to be considered as predominantly neutral in expression. This was expected as it may interfere with the Microsoft API's ability to distinguish features necessary to determine a predominant emotion on a face.

4 Discussion

For these purposes an API that would provide informative and accurate results was the top priority. This required it to return emotion information for a reasonable amount of faces, and that the emotional information it returned was reliable.

We derived emotional information for faces captured in broadcast streams of tennis matches. Three APIs were accessed through the REST system, we were able to collect the emotional information it provided for the 1319 faces in the set. However, responses did not contain emotional information for all of the provided faces. Microsoft and Skybiometry reponded with many NA results, and Google responded with many faces that were "very unlikely" to be displaying emotions.

It has been shown that Goggle returned emotional information for 910 player's faces, the highest amount of emotion results for the faces provided to three APIs. This may have been due to the faces being used already having been detected by Google previously. It was unexpected that Skybiometry would only be able to produce emotional results for just under half of the images presented to it.

It was presumed that the algorithms may have produced emotional information for different faces based on their performances in a previous study where the focus was facial detection. This was apparent as 364 faces were found by all APIs, 113 by only Microsoft and Google as well as 53 found by Skybiometry and Google. However the amount of faces found by only Google, 380, was higher than anticipated.

As there was already disagreement between the APIs in finding faces we were inclined to believe that the approaches to classifying faces, and therefore emotions, may be different. While more facial emotion results were found by Google, the majority of the emotion proportions after transformations were still below 0.5 for the predominant emotions. This amount of 'Very Unlikely' responses to emotions being present on a faces was surprising. We did not expect that Google would be reserved in aknowledging the prescence of emotion. Microsoft provided 499 emotion results for player's faces. These results had the most emotion categories, we considered this to be a benefit to us as it may allow an emotion to be directly identified beyond the four Google categories. After the calculation of a predominant emotion we were surprised to find 0.0791209 of the faces Google returned results for were non-neutral. This small amount of emotional information would not be helpful in extracting a player's emotional state and tracking changes throughout a match. Given that both Google and Microsoft were not inclined to allocate high proportion levels to the array of emotions it was worrisome that Skybiometry would return high confidence values for the emotion being present on the face. However, Microsoft provided many results where the amount attributed to an emotion was zero, this was accompanied by high values of neutral being counted in these faces, as shown in the emotion histograms of Google and Microsoft. This is not unreasonable as there are factors that may have influenced the neutral results, such as the faces passed to the software being sub optimal for recognition.

We suspected that the amount of faces classified as neutral would mean that only extremely obvious facial expressions would be considered predominant emotions. Therefore the intersections of the predominant emotion classifications were expected to be quite high, and the strong disagreement between the APIs was unexpected. It was surprising that no faces in the set had the same predominant emotion result from all three softwares, though there were a small portion of faces classified as the same predominant emotion by Microsoft and Skybiometry. This enforces the need for further validation to be conducted on these APIs. This blatant disagreement led to confusion in a choice of API should further work using this method be pursued.

The research aim was to develop methods to collect accurate information about the facial expressions of elite tennis athletes during matchplay. We cannot establish, without further validation, that three currently available APIs are not capable of accurate detection of player's emotions when analyzing the broadcast video matches. Yet, the alternative approach to answer the question of whether a player's mental state correlates with their on court performance is worth pursuing. This would allow the coaches and players evidence of the player's ability to focus and perform mentally in elite tennis matches. If there are correlations between the engagement or concentration levels and their performance during a match this would help players recognise how their mental state impacts them.

Given the low amounts of emotion classifications we considered that the FACS method may not be the most effective for our purposes as the amount of neutral results may detract from any attempt to discover correlations between emotions and performance. In this elite tennis setting we may gain more informative insights into the mental state of a player from searching for the levels of concentration and engagement rather than emotions. We should also recognise that the APIs are attempting to perform with sub optimal images and any further work may also be limited if the sample will be the size used for this set. To make inferences about a player and their mental state the set would need to be expanded to consider individuals in more depth. If the solution is able to be trained for individuals this would help prevent facial features influencing FACS results.

5 Possible Application

From this point we will give an example of a use for this emotional data in a tennis setting. Should an API prove to be returning valid emotion categorizations this information would help to understand a player's emotional variability on the court. It could then be used help coaches training players to control what is physically visible as to not advantage their opponent by giving them an indication to a player's emotional state.

5.1 Player Variablity

We consider the subsets corresponding to individual players. This allowed a visual inspection of a players emotional variability according to each software.

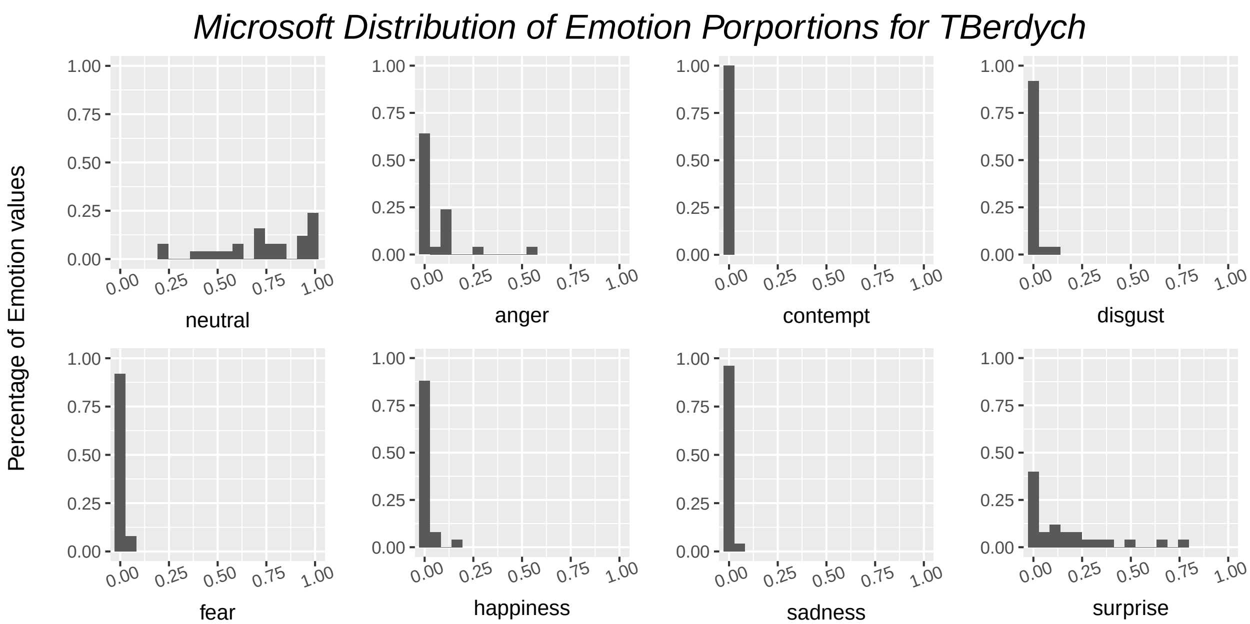


Figure 7:

Tomas Berdych had the most faces recognised, 25 of the 943 player faces. Figure 7 shows a similar distribution to that seen in the full Microsoft Distribution in 5. There are no values of neutral that occur for more than 0.25 of the total set of images of Berdych's face however the neutral values are all above 0.10 for Berdych. Similarly to the full distribution, the distribution that departs the furtherst from the distribution of the whole sample is for anger, where there are proportionally less faces with an anger value of zero, with around .25 of the faces having an anger value of 0.1.

Assumptions about the algorithm's performance cannot be made based on the results produced when considering Alize Cornet's expressions. Microsoft was only able to produce emotional information for 6 of the 12 images captured of Cornet's face. However it does show levels of surprise that are quite high, especially when compared to the full distribution.

6 Future Work

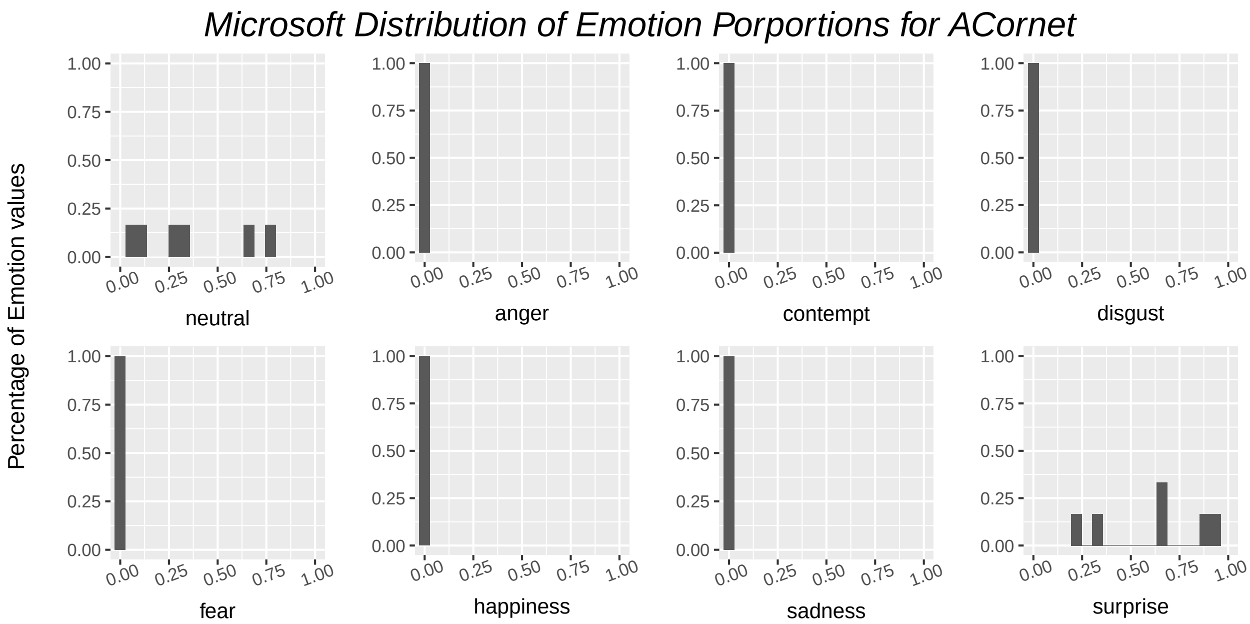


Figure 8:

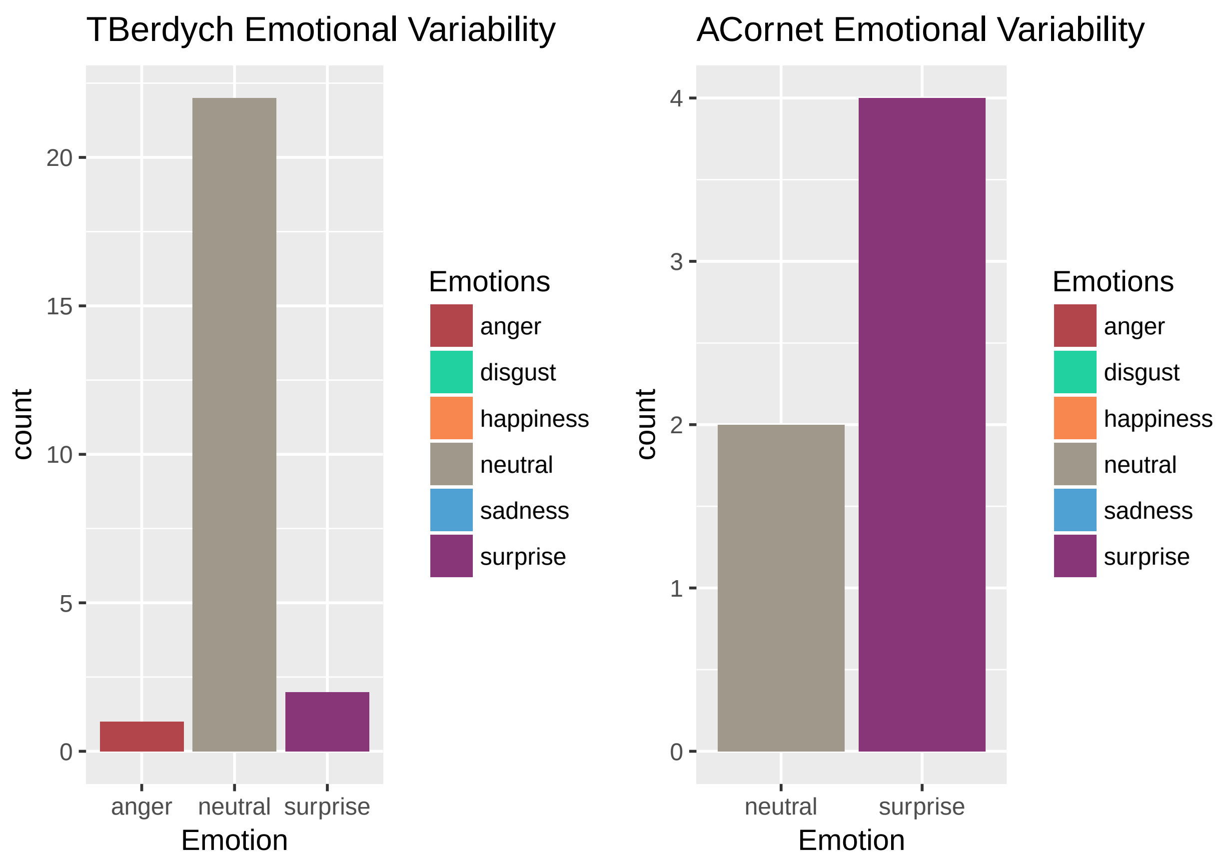


Figure 9: