

BEX3012 Project Report Detecting Facial Expressions in Professional Tennis Matches

Stephanie Kobakian

28 March 2017

Contents

1	Introduction	2
2	Materials	3
2.1	Images	3
2.2	Software APIs	3
3	Procedure	5
3.1	Process Images	5
3.2	Process API results	5
4	Analysis	8
4.1	Selecting an API	14
4.2	Possible Application	15
4.3	Player Variability	15

1 Introduction

Facial Recognition is beginning to be explored in sports environments, this presents quality issues.
helpful for applications

2 Materials

2.1 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 a combination of multiple camera feeds, these angles vary depending on the court the match was played on.



Figure 1: This image was derived from a match played between JMillman and DSchwartzmann in the first Round of the 2016 Australian Open.

2.2 Software APIs

Three emotion recognition APIS were considered. They were chosen for their accessibility and online reviews of their performance, and these are well recognised options for Facial recognition currently available.

Table 1: This details the capabilities we considered important in recognising 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

As can be seen 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 be 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.

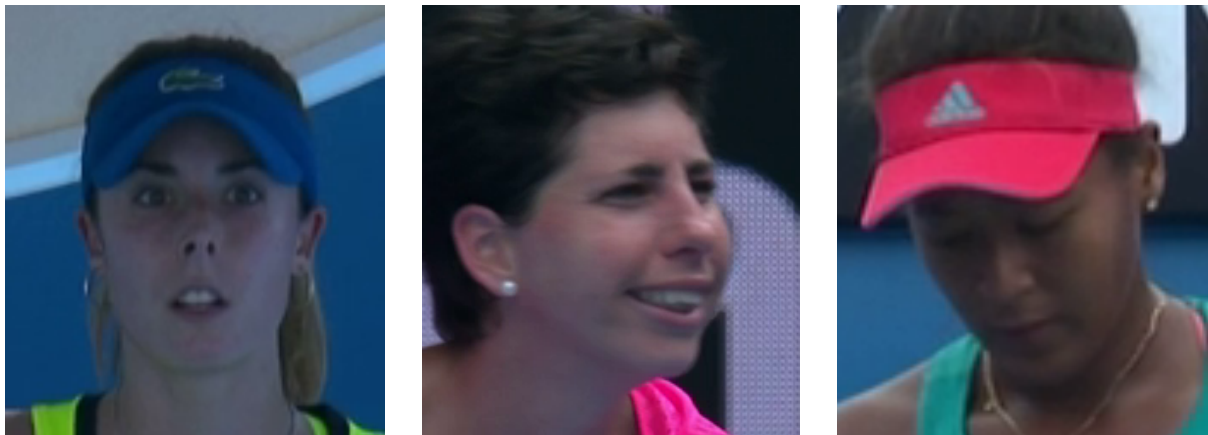
The number of emotions shown in the results were not congruent. These differences made it necessary to process the outputs to provide the most comparable results for analysis. Microsoft provided the most emotion options, contempt being the addition that was not provided by Skybiometry. Google provided the least, only providing four different emotions.

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, whereas Skybiometry results in a Confidence value for the emotion occurring in the specified face in values from 0 to 100.

3 Procedure

3.1 Process Images

To create a set of individual faces for the APIs to consider we derived small images that were cropped subsets of the 450x800px images. The crop was performed to create a new image with a unique name that focused on the area marked as a face by Google¹. A border to frame the face was also included for visual appeal. These 1319 new images, of varying sizes, were hosted on Google Drive as individual images.



Software	Emotion	Software	Emotion	Software	Emotion
Microsoft	surprise	Microsoft	happiness	Microsoft	NA
Skybiometry	surprise	Skybiometry	fear	Skybiometry	NA
Google	neutral	Google	joy	Google	neutral

Figure 2: Three Faces in the image set, after extracting faces from the full broadcast video stills, and the dominant emotion each software recognised for them.

3.2 Process API results

As mentioned previously, the results extracted for each face differed across the APIs this made it necessary to produce a comparable set of results for analysis.

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

¹based on Google's bounding box coordinates

values were not returned for a certain face, a row of NAs will be associated with the unique face ID in the data set.

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

anger	contempt	disgust	fear	happiness	neutral	sadness	surprise
0.001504	0.000631	0.002371	0.003267	3e-05	0.716105	0.004738	0.271353

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 100’² representing the confidence of the emotion being present on a player’s face.

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

neutral	anger	disgust	fear	happiness	sadness	surprise
0	21	41	0	0	5	76

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.

joy	sorrow	anger	surprise
VERY_UNLIKELY	UNLIKELY	VERY_UNLIKELY	UNLIKELY

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
---------	-------	---------	------	-----------	---------	----------

²<https://skybiometry.com/documentation/#document-21>

0	0.146853	0.286713	0	0	0.034965	0.531469
---	----------	----------	---	---	----------	----------

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:

$$neutralConf_i = 100 - max(joyConf_i, sorrowConf_i, angerConf_i, surpriseConf_i)$$

The process to derive the numeric values for Google followed the process undertaken for Skybiometry values. Where the individual emotion confidences were divided by the sum of the emotion confidence values for each face. This resulted in congruent emotion data.

Table 7: The Google Vision API provided four likelihoods, one for each emotion.

joy	sorrow	anger	surprise	neutral
0.0666667	0.2	0.0666667	0.2	0.4666667

4 Analysis

Each API returned a different amount of non null results. 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 results³.

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 382 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

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.

³As 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.

4.0.1 Emotion Proportion Distributions

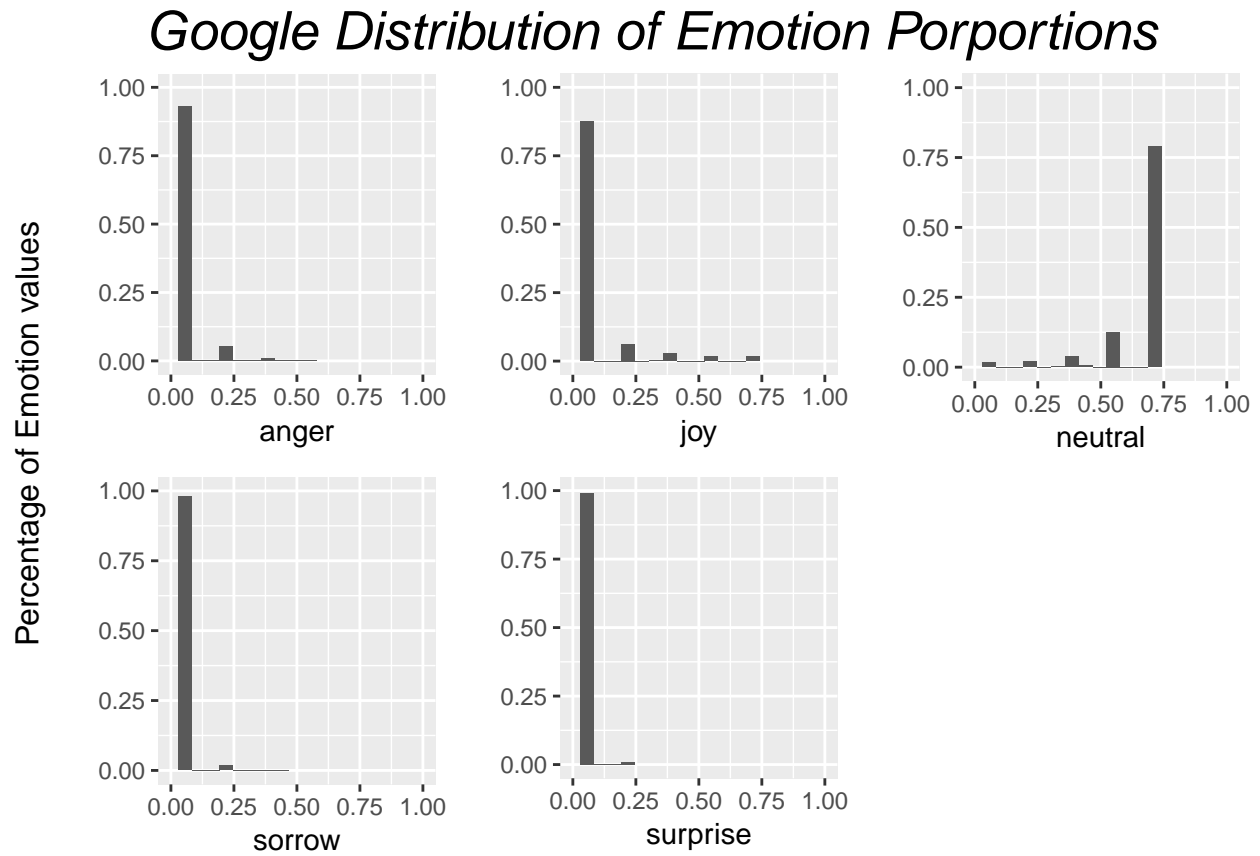


Figure 3: It can be seen that there are significant proportions of faces where the non neutral emotions occur close to zero. This is balanced by the large percentage, over 75%, where the neutral value of just under .75 occurred. The faces used showed higher levels of Joy than any other non neutral emotion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

Skybiometry Distribution of Emotion Porportions



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.

Microsoft Distribution of Emotion Porportions

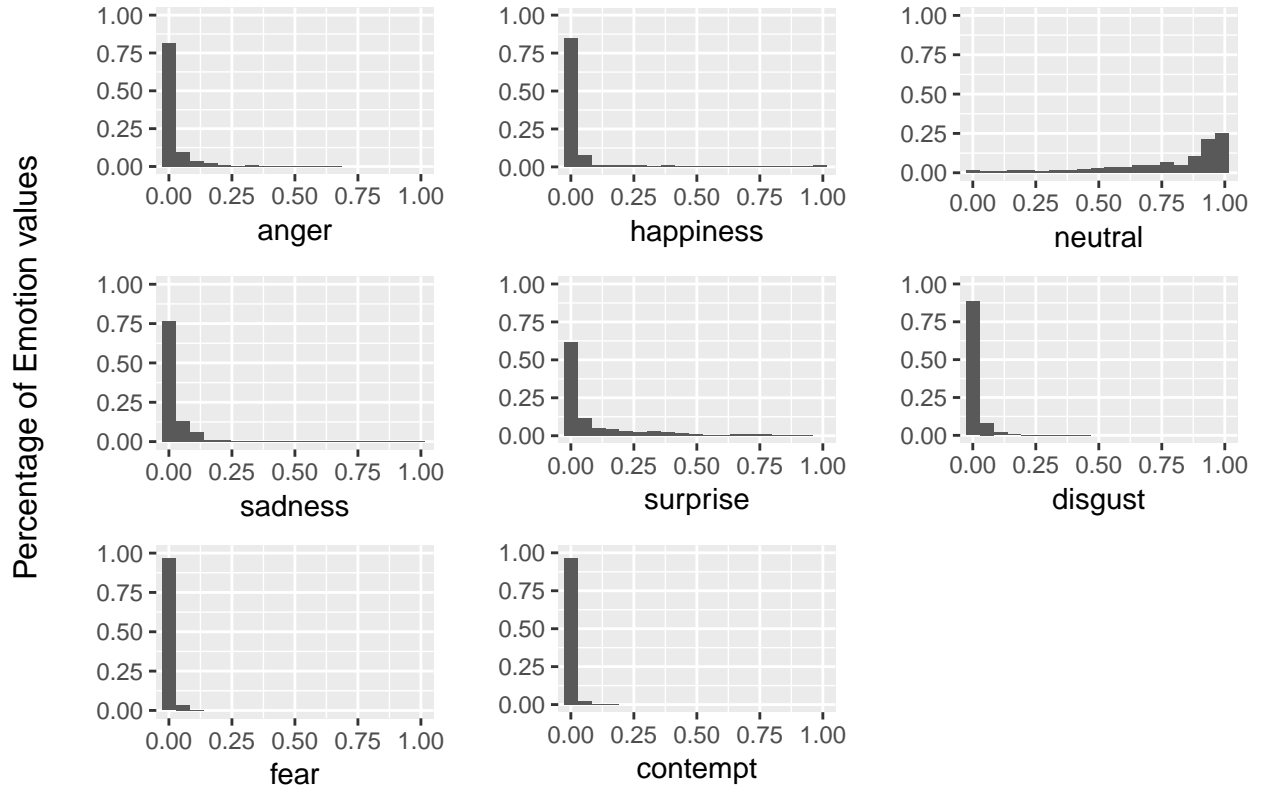


Figure 5: There were many faces that returned high neutral values. This is matched by the high percentages of faces where the proportion levels were at, or close to, zero. While the happiness emotion was most varied, with values between 0 and 1, over 40% of faces returned surprise levels above zero.

4.0.2 Predominant Emotions

To contrast the emotion results for certain faces a ‘predominant’ emotion was designated to each face. Where the emotion with the maximum proportion value was deemed the ‘predominant’ emotion for a particular face.

This resulted in one of seven emotions for Google, one of eight for Microsoft, and one of five for Google. This also required the assumption that joy is equivalent to happiness, and sorrow to sadness.

Table 9: This shows the agreement between the emotions recognised by Microsoft and Skybiometry

		Skybiometry	anger	disgust	fear	happiness	neutral	sadness	surprise
Google	anger	Microsoft							
		anger	0	0	0	0	0	0	0
		disgust	0	0	0	0	0	0	0
		happiness	0	0	0	0	0	0	0
		neutral	5	4	1	0	0	2	0
		sadness	0	0	0	0	0	0	0
joy	surprise	surprise	0	0	0	0	0	0	0
		anger	0	0	0	0	0	0	0
		disgust	1	1	0	0	0	0	0
		happiness	0	3	3	0	0	1	1
		neutral	3	3	0	1	0	0	0
		sadness	0	0	0	0	0	0	0
neutral	surprise	surprise	0	0	0	0	0	0	0
		anger	1	0	1	0	0	0	1
		disgust	0	0	0	0	0	0	0
		happiness	2	0	0	0	0	2	0
		neutral	82	62	46	0	3	37	72
		sadness	1	1	0	0	0	1	0
sorrow	surprise	surprise	1	3	3	0	0	2	12
		anger	0	0	0	0	0	0	0
		disgust	0	0	0	0	0	0	0
		happiness	0	0	0	0	0	0	0
		neutral	2	0	0	0	0	0	0
		sadness	0	0	0	0	0	0	0
		surprise	0	0	0	0	0	0	0

Skybiometry emotion recognition’s are shown in the rows, and Microsoft’s in the columns. There were many instances where Skybiometry allocated an emotion as the predominant expression while Google and Microsoft designated it as neutral, this was expected as only 3 faces were categorised as neutral by Skybiometry. These faces were categorised as neutral by both Google and Microsoft.

As neutral was a prominent category for Microsoft and Google, we considered where neutral was the dominant emotion and grouped the remaining emotions into a non-neutral category.

The following Figure 6 displays a Scatter Plot Matrix that captures the pairwise variability in the Neutral proportion levels across APIs.

Table 10: 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
Google	neutral		3	299
	non-neutral		0	31
Microsoft	neutral		0	21
	non-neutral		0	10

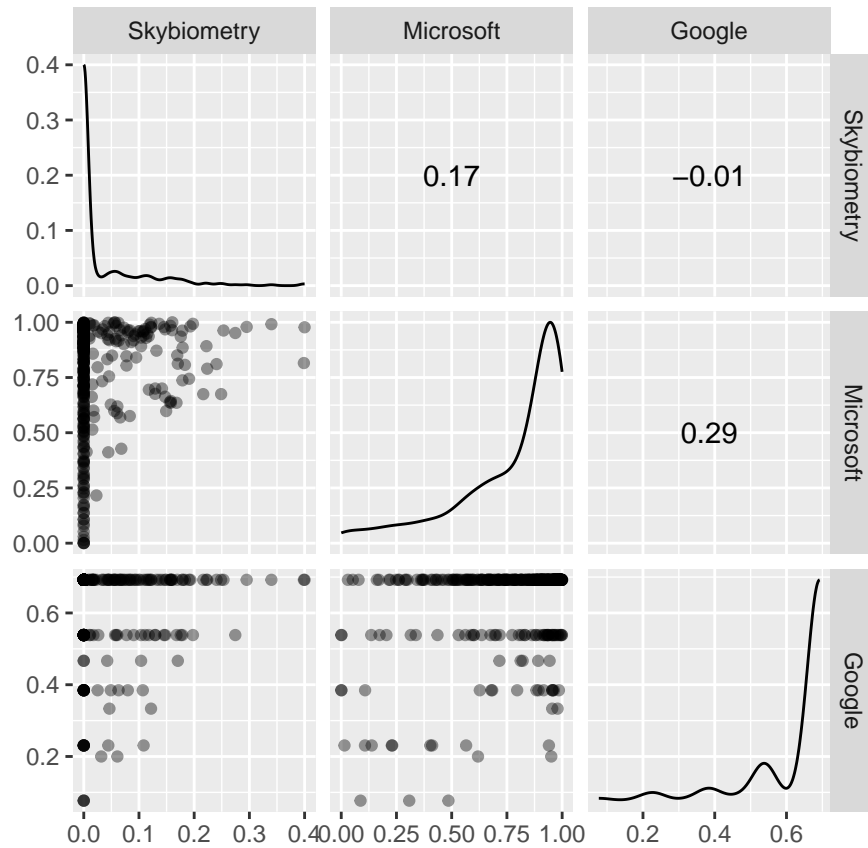


Figure 6: The Scatter Plot Matrix shows the variability in the Neutral proportion levels. The -0.01 correlation value for Google and Skybiometry show that they do not behave either in a similar way, or inversely. They are uncorrelated. Some high values of neutrality for Google are matched to some high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

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 differing attribute values and the neutral results.

4.1 Selecting an API

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

It has been shown through the report that Goggle returned emotional information for 910 player's faces, the highest amount of emotion results for the faces provided to three APIs. However, in Table 6 it is seen that only 31 of these faces gave results that considered the faces to be premoninantly a non-neutral expression. This limited resulting information is not helpful in extracting a player's emotional state. Figure 3 shows that anger, joy, sorrow and surprise often had a level of 0.10 for faces. This is equivalent to the original output of it being 'Very Unlikely' that the emotion appeared on the given face. Due to these results, Google will not be appropriate for further exploration.

Microsoft provided 499 emotion results for player's faces, these numeric results shown in Table 2 were in an optimal format for analyses. These results also had the most variability is possible emotions, this could be seen as a positive as it may allow an emotion to be directly identified in comparison to the assumptions needed to be applied to Google's results in Table 4.

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 Figure ???. This is not unreasonable as there are many factors that may have influenced the neutral results we were given, such as the players having strong control over their facial expressions or the faces passed to the software being sub optimal for recognition.

The results from Microsoft were still deemed to be preferable to those of Microsoft due to the limits on it's API impeding on any future use of the API for our purposes. While more facial emotion results were gained, majority of the emotion proportions after tranformations were still below 0.5 for the predominant emotions.

The conclusion of these comparisons points to Microsoft as a good API that provides a useful output. Its limits do not impose as heavily as Skybiometry's do, and it provided many of the same neutral results that Google did, showing that it may be image, rather than API, reasons for returning a neutral dominant emotion result.

4.2 Possible Application

From this point we will give an example of an application of this emotional data in a tennis setting. This information would help to understand a player's emotional variability on the court, and may help in coaches training players to control what is physically visible as to not advantage their opponent by giving them an indication to a players emotional state.

4.3 Player Variability

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

