Going Inside the Inner Game: Predicting the Emotions of Professional Tennis Players from Match Broadcasts

Stephanie Kovalchik and Machar Reid
Game Insight Group, Tennis Australia
Email: skovalchik@tennis.com.au

Other Sports (Tennis) ID 5564

Abstract

The mental side of the game has been one of the most elusive aspects of performance analysis in tennis. We present a framework for predicting seven emotional states relevant to sport (‘anxiety’, ‘anger’, ‘annoyance’, ‘dejection’, ‘elation’, ‘focus’, and ‘fired up’) from the observed facial expressions of players in match broadcasts. Our methodology applies pre-trained models to extract two feature sets: predicted emotions in the Facial Action Coding System and 17 facial action units. Multiple prediction approaches were trained and tested using these features and a labeled dataset of 1,700 facial images of professional male and female tennis players extracted from 505 match videos. We applied the prediction models to establish emotional profiles for the ‘Big 4’ (Roger Federer, Rafael Nadal, Andy Murray, and Novak Djokovic) at the 2017 Australian Open. Rafael Nadal exhibited the most ‘anxiety’ of the four players (32%, 95% CI 29 to 35%), while Roger Federer was the only player whose predominant state was ‘neutral’ (24%, 95% CI 21 to 27%). When the predicted emotions were associated with point outcomes, we found that all of the Big 4 except for Roger Federer showed significant emotional reactions to the outcomes of points. Further, several emotional states of Rafael Nadal and Novak Djokovic were significantly predictive of their chances of winning the next point. Our framework for extracting emotional data from single-camera video in professional tennis shows the feasibility of bringing the quantitative study of the inner game into sports performance analysis.

1. Introduction

Throughout its history, tennis has often been described as a ‘mental game’. The individual nature of the sport and the fact that 80% of the time in a tennis match is spent preparing to play (Kovacs 2006) have both contributed to a heavy emphasis on player mentality in tennis coaching and commentary. Indeed, the role of mentality is a central topic in some of the most popular training texts in the sport, including Loehr’s ‘16-second cure’ (Loehr 1990) and Gallwey’s The Inner Game (Gallwey 2008). Although these and other popular theories about the mental game have a lot of traction in the sport, few if any of these theories have been rigorously tested with real data.

Theoretical work about the role of emotions in sport (EIS) is extensive. A common theme among the myriad proposed theories is the belief that the presence and intensity of emotions in competition can influence player performance in both negative and positive ways (Lazarus 2000; Hanin 2012). Testing this fundamental idea about the influence of emotion on sport performance requires accurate measurement of emotion in competition. Yet, the measurement of EIS has largely
relied on qualitative approaches, such as video-assisted interviews or survey instruments (Anthony, Gucciardi, and Gordon 2016; Gucciardi 2017), several of which have been developed specifically for the assessment of aspects of athlete mentality (Gucciardi 2012). While qualitative approaches can be useful for the study of player perceptions about their emotional states, they do not provide an objective measure of in-competition emotion, and attempts to use them in this capacity raise multiple validity concerns including recall bias, information bias, and emotion misclassification (Murray et al. 2004).

There has been very limited quantitative research on EIS. In tennis, several studies have attempted to measure psychological effects in sport through indirect means based on observed changes in performance with changes in game pressure (Klaassen and Magnus 2001; Kovalchik and Ingram 2016). Without some measure of the mental state of players in these situations, however, it is not possible to disentangle the psychological nature of these effects from other possible causes, like strategic adaptation to game situation. Also, the lack of more direct measurement of their psychological states limits our understanding about the mentalities of elite athletes.

Over the past two decades, there has been extraordinary progress using machine learning to perform a variety of facial recognition tasks with video and image data (Sariyanidi, Gunes, and Cavallaro 2015). Propelled by crowd-source competitions (Dhall et al. 2013) and the establishment of large public datasets of facial images and emotion labels (Bainbridge, Isola, and Oliva 2013), there are now well-established tools for complex facial recognition tasks and a growing body of methods applying machine-detected facial features to the prediction of emotions (Littlewort et al. 2011; Baltrušaitis, Robinson, and Morency 2016). Given the proliferation of video and image in modern sport, the lack of image-based expression detection is a major shortcoming of current sports performance research.

There are numerous examples of how machine learning has been used to develop models to predict emotions from images (Kahou et al. 2013; Khorrami, Le Paine, and Huang 2015; Rajesh and Naveenkumar 2017). Most methods that have been considered have been limited to the prediction of the basic emotion states included in the Facial Action Coding System (FACS) (Ekman and Friesen, 1978): ‘anger’, ‘fear’, ‘sadness’, ‘happiness’, ‘surprise’, and ‘disgust’. Although FACS emotions are expected to be present in the sporting context, their relevance may be limited as they, according to qualitative work in sports psychology (Laborde, Raab, and Dosseville 2013; A. Lane et al. 2012), do not encapsulate the primary emotions experienced in competition. Thus, while prior emotion prediction work shows the feasibility of extracting rich facial expression data from single-camera video, the accuracy of these ‘off-the-shelf’ methods in the sports setting is unknown and the relevance of their emotional categories is likely limited.

The aim of the present study was to develop a methodology for predicting emotions in sport from single-camera broadcast video of professional tennis matches. The sport-relevant emotions that were included in the study consisted of the following seven emotions: ‘anger’, ‘annoyance’, ‘anxiety’, ‘dejection’, ‘elation’, ‘focus’, and ‘fired up’. Our model development process involved the construction of a labeled image dataset of the sport-relevant emotions, feature extraction of standard FACS emotions and facial action units (FAUs), and training and evaluation of multiple supervised machine learning approaches. In what follows, we detail the model development and investigate the out-of-sample prediction performance in our test set of player images. To illustrate the usefulness of these models, we apply them to matches at the 2017 Australian Open to characterize the emotional profiles of the ‘Big 4’—Novak Djokovic, Roger Federer, Andy Murray, and Rafael Nadal—and evaluate the link between their observed emotions and their performance.
2. Datasets

Three image datasets were compiled for the different stages of model development and evaluation: (1) FACS feature, (2) tennis player in-match and (3) ‘Big 4’ datasets. The FACS feature dataset included two public image datasets: the Facial Expression Recognition Challenge 2013 data (Goodfellow et al. 2013) and the affect-labeled 10k US Adult Faces Database (Bainbridge, Isola, and Oliva 2013). Each dataset included cropped images of male and female human faces and each image had an indicator for the predominant FACS emotion shown (including a ‘neutral’ category) according to a consensus vote across multiple independent raters. All images where FAUs could be detected with the OpenFace toolkit (Baltrušaitis, Robinson, and Morency 2016) were retained, resulting in a total set of 7,952 FACS-labeled images.

The construction of the tennis player in-match dataset began by extracting frames from high-definition broadcast video of men’s and women’s matches played at the 2014 to 2016 Australian Opens. Using the ffmpeg video conversion tool (Tomar 2006), frames were extracted at a rate of 1 frame per second. Facial detection and cropping was performed with OpenFace software. The images were reviewed by the authors to ensure they met the following inclusion criteria: an unobstructed player face in-between play. From this set, a random sample of 1,700 images were selected for labeling of the seven EIS: ‘anger’, ‘annoyance’, ‘anxiety’, ‘dejection’, ‘elation’, ‘focus’, and ‘fired up’. These emotions were selected from the Sport Emotion Questionnaire (Jones et al. 2005), which is an established psychometric instrument in sport psychology.

EIS labeling was performed using the Amazon Mechanical Turk service (Paolacci and Chandler 2014). The designed task for each worker included a single image and a request to rate the perceived intensity of a specific EIS on a scale from 0 to 10 (10 being the most intense). Five independent ratings were obtained for each EIS and image. Only qualified workers were eligible to complete the task. The qualification criteria was proficiency in English, 95% or greater historical approval rate, and successful completion of an EIS assessment that required EIS classification on 5 images within +/- 1 point of the labels of the authors. The median of the five ratings was the final EIS intensity assigned to each image.

For our application study, we used the same methods as described above to extract face-detectable images at a rate of 1 frame per second for match videos of matches played by Novak Djokovic, Roger Federer, Andy Murray, and Rafael Nadal at the 2017 Australian Open, including two matches per player. Detected facial images were reviewed by the authors and images of faces other than the ‘Big 4’, obstructed faces, or frames during play were excluded. A total of 3,420 images (n = 972 for Djokovic, n = 908 for Federer, n = 509 for Murray, and n = 1,031 for Nadal) were included in the Big 4 dataset.

3. Feature Extraction

A major part of the EIS prediction model development, as diagrammed in Figure 1, was the extraction of two feature sets: (1) FAUs and (2) predicted FACS emotions. The following
subsections describe the components of each feature set and the methods used to extract them from an image file.

3.1. Facial Action Units

OpenFace is open-source software for facial analysis written with python and torch (Baltrušaitis, Robinson, and Morency 2016). The feature of the toolkit used in the present study was its facial action unit detection. This feature provides presence and intensity information on 17 FAUs (e.g. jaw drop, blink, etc.) in real-time. The core of the methodology is the use of Constrained Local Neural Fields for geometrical features and Histograms of Oriented Gradients for appearance features, which are then combined to predict each action unit using a linear support vector machine (SVM). The tool was trained on multiple public datasets and has been shown to outperform standard benchmarks, especially with dynamic inputs. Further details can be found in Baltrušaitis, Mahmoud, and Robinson (2015).

3.2. Basic FACS Predictions

For each of the images in the FACS feature dataset, we extracted the presence and intensities of FAUs using OpenFace software. Centered and scaled FAU features were trained to predict each of the six basic FACS emotions. Training and testing was performed separately per emotion category. In each case, images for a given emotion category were included if the image either had the given emotion as its predominant

<table>
<thead>
<tr>
<th>FACS Emotion</th>
<th>Accuracy (%)</th>
<th>AUC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>71</td>
<td>0.67 (0.63 - 0.71)</td>
</tr>
<tr>
<td>Disgust</td>
<td>94</td>
<td>0.68 (0.60 - 0.75)</td>
</tr>
<tr>
<td>Fear</td>
<td>68</td>
<td>0.66 (0.62 - 0.70)</td>
</tr>
<tr>
<td>Happy</td>
<td>83</td>
<td>0.85 (0.82 - 0.87)</td>
</tr>
<tr>
<td>Sad</td>
<td>63</td>
<td>0.64 (0.60 - 0.67)</td>
</tr>
<tr>
<td>Surprise</td>
<td>80</td>
<td>0.78 (0.74 - 0.81)</td>
</tr>
</tbody>
</table>

Table 1. Prediction accuracy and AUC in test data for binary classification of basic FACS emotions
emotion or was neutral. The emotion outcome was treated as a binary classification problem and was trained with a random sample of 70% of the applicable sample and tested on the remaining 30%. Among a range of standard machine learning approaches, we found the model accuracies in the test sample was generally highest with radial SVMs, yielding the results shown in Table 1. Consequently, these were the models used to derive basic FACS features for the subsequent EIS image datasets.

3.3. Feature Associations

As it can be difficult to interpret the relative importance of features in machine learning models, we first performed linear regression analysis to understand the general direction and strength of associations between the feature sets and labeled EIS intensities in the training dataset (described below). The t-statistic for each feature coefficient, which is scaled to the covariate’s standard error, was the effect size used for making comparisons of the relative importance across features. These did not influence the choice of feature inclusion but were used purely for descriptive purposes and improved interpretation of the prediction results.

![Figure 2. Effect sizes (t-statistic) for the association of input features and emotion of sport intensities based on a linear regression in the tennis in-match training sample. Input features include predictions for the six standard Facial Action Coding System (FACS) emotions, where neutral is the reference category, and detected intensities for 17 Facial Action Units (FAUs).](image)

Most of the EIS emotions showed strong associations with one or more of the FACS and FAU features (Figure 2). Among the FACS features, two of the strongest negative associations were found between FACS-predicted ‘anger’ and EIS ‘dejection’ (t=-5.8) and ‘surprise’ and EIS ‘anger’ (t=-5.3). Two of the strongest positive associations were observed with EIS ‘anxiety’, one association
with FACS-predicted ‘fear’ (t=7.5) and the other with ‘surprise’ (t=6.8). Even stronger associations were found with FAUs. The action of ‘lips part’, for example, had a strong negative association with EIS ‘dejection’ (t=-7.1) but a strong positive association with EIS ‘anxiety’ (t=10.1). Similarly, a ‘cheek raiser’ action was strongly negatively associated with ‘anger’ (t=-7.5) but highly positively associated with ‘elation’ (t=6.5). The adjusted R-squared statistics from the linear regression models of the combined feature sets suggested moderate to strong explanatory value across the EIS: ‘focus’, 23%; ‘anxiety’, 23%; ‘annoyance’, 24%; ‘fired up’ 24%; ‘dejection’, 36%; ‘anger’, 49%; and ‘elation’, 71%.

4. Model Training and Performance

4.1. Methods

For the prediction model development, the in-match tennis dataset was divided into training and test samples using a 70%/30% split. Balance across the 0-10 intensity classes was achieved by up-sampling to the largest class. Two major sets of specifications determined the scope of model training: (1) the included feature sets and (2) the prediction method. Three feature sets were considered: standard FACS predictions, detected presence and intensities of FAUs, and both feature sets combined. All features were standardized prior to model fitting.

Fifteen different machine learning models from the R caret package were evaluated during training (Kuhn 2008). The selected methods span the major categories of machine learning approaches: support vector machines, discriminant analysis, neural networks, boosting, bagging, and regularized regression. For each tuning parameter of a given model, multiple values over a parameter grid were considered. Within each feature set, the choice of method and tuning settings were based on the approach that minimized the root mean squared error (RMSE) of predicted intensities in a 5-fold cross-validation of the training data.

The selected model for each feature set was then applied to the test sample and the following prediction measures were collected: RMSE and the F1 score, a combined summary of precision and recall. Owing to the multiclass nature of the EIS predictions, the RMSE metric was used for selecting the preferred model.

4.2. Predictive Performance

The prediction method that showed the best performance overall was an SVM model with a radial cost function (svmRadialCost), which was the selected approach for four of the seven emotions (Table 2). A different method was found to give the best performance for each of the three remaining EIS, including a boosted generalized linear regression approach (glmboost), k nearest neighbors (kknn), and a penalized regression approach (lasso).

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Method</th>
<th>FACS</th>
<th>FAUs</th>
<th>FACS + FAUs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>F1</td>
<td>RMSE</td>
</tr>
<tr>
<td>Anger</td>
<td>glmboost</td>
<td>2.83</td>
<td>0.17</td>
<td>2.91</td>
</tr>
<tr>
<td>Annoyance</td>
<td>svmRadialCost</td>
<td>3.48</td>
<td>0.18</td>
<td>3.33</td>
</tr>
<tr>
<td>Anxiety</td>
<td>kknn</td>
<td>4.01</td>
<td>0.10</td>
<td>3.90</td>
</tr>
<tr>
<td>Dejection</td>
<td>lasso</td>
<td>3.93</td>
<td>0.09</td>
<td>4.02</td>
</tr>
</tbody>
</table>
Table 2. Selected EIS prediction models and performance across Facial Action Coding System (FACS) and Facial Action Units (FAUs) feature sets

<table>
<thead>
<tr>
<th>Model</th>
<th>Loss 1</th>
<th>Loss 2</th>
<th>Loss 3</th>
<th>Loss 4</th>
<th>Loss 5</th>
<th>Loss 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elation</td>
<td>2.91</td>
<td>0.21</td>
<td>2.87</td>
<td>0.31</td>
<td>2.53</td>
<td>0.25</td>
</tr>
<tr>
<td>Focus</td>
<td>4.01</td>
<td>0.13</td>
<td>3.91</td>
<td>0.14</td>
<td>3.90</td>
<td>0.21</td>
</tr>
<tr>
<td>Fired up</td>
<td>3.23</td>
<td>0.13</td>
<td>3.03</td>
<td>0.12</td>
<td>2.97</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Within each prediction approach, we consistently found an improvement in performance with the inclusion of both the FACS and FAU feature sets as shown by the reduction in RMSE (Table 2). Across all of the EIS and the combined feature sets, the RMSE for the emotion intensity in the test data ranged from a low of 2.53 (‘elation’) and a high of 3.90 (‘focus’). Treating the intensity score as a multiclass outcome, the F1-scores ranged from a low of 0.15 (‘annoyance’ and ‘fired up’) to a high of 0.25 (‘elation’).

5. Practical Application

5.1. Emotional Profiles

A primary purpose of the EIS predictive models is to enhance our understanding of player emotions in competition. To demonstrate this use, we applied the models to facial images of the Big 4 at the 2017 Australian Open. Each image yields a predicted intensity for every EIS. A ‘predominant’ emotion for a given image was defined as the emotion with the highest predicted intensity, with neutral being assigned in cases where no EIS had an intensity of 5 or more.

The frequency of a player’s predominant emotions provides a profile of their emotional baseline at an event. Figure 3 summarizes the observed profiles among the Big 4 and shows some interesting differences among this elite group of players. At a frequency of 32% (95% CI 29 to 35%), ‘anxiety’ was the most predominant emotion observed for Rafael Nadal and the highest single emotion for all players. Roger Federer, on the other hand, was the only player of the Big 4 that was most often ‘neutral’ (24%, 95% CI 21 to 27%) or ‘focused’ (23%, 95% CI 20 to 26%).

High levels of anxiety were also observed for Andy Murray (28%, 95% CI 24 to 32%) and Novak Djokovic (23%, 95% CI 21 to 26%). Several other emotional states were more common for Djokovic, however, than any other player. These included ‘fired up’ (18%, 95% CI 15 to 20%), ‘dejection’ (17%, 95% CI 14 to 19%), and ‘anger’ (11%, 9 to 13%). The nearly equal representation across several EIS suggest that Djokovic was the most expressive player out of this group at the 2017 Australian Open.

Although ‘elation’ and ‘annoyance’ were two of the rarest emotions among the EIS, they were expressed most often by Andy Murray (both 6%, 95% CI 4 to 8%). When it comes to emotions on court, Murray is considered to be one of the more mercurial players on tour and the high relative frequency of these two seemingly opposing emotions support this conclusion.
Figure 3. Predominant emotions among Big 4 at the 2017 Australian Open. Each point shows the frequency (as a percentage) that the emotion in sport (EIS) was the predominant emotion. Lines indicate the 95% confidence interval for the frequency. The player with the highest frequency for a given EIS is highlighted in orange.

### 5.2. Performance Link

Two common beliefs in tennis are that 1) players respond emotionally to how they perform and 2) emotional reactions influence their future performance in a match. We tested both of these ideas by merging data on the game context with each facial image in the Big 4 dataset. The contextual information included information about the score, the winner of the point, the server of the point, and the point’s importance.
To assess emotional reactions to point outcomes (what we refer to as ‘reactive associations’), we estimated the association between the EIS predicted intensities and the outcome of winning the most recently completed point. The percentage increases (decreases) shown reflect the expected change in the odds of winning the point with one standard deviation (SD) increase in the EIS intensity. A percent increase in the odds ratio for a specific EIS would indicate an expected increase in the EIS intensity after winning a point, whereas a percent decrease in the odds would indicate an expected increase in the EIS intensity after losing a point. In all of the regression models, the EIS intensities were standardized and additional adjustment variables were included for a service point and point importance (Morris 1977).

![Emotion and Winning Point](image)

*Figure 4. Percentage change in point winning odds associated with standard deviation increase in emotions in sport (EIS) after the point was played (‘Reactive’) and before the next point was played.*
Evidence of significant emotional reactions were found for three of the Big 4 players at the 2017 Australian Open (left column, Figure 4). Andy Murray's reaction profile showed the largest effects of any of the players, with large increases in 'elation' (26%, 95% CI 5 to 48%) and 'focus' (41%, 95% CI 20 to 70%) and a notable reduction in 'dejection' (-18%, 95% CI -27 to -1%) after winning a point, emotional swings that Murray has himself recognized as characteristic of his on-court demeanor. Rafael Nadal showed a similar pattern to Murray in how his predicted 'focus' (12, 95% CI 2-26%) and 'dejection' (-14, 95% CI -22 to -4%) changed with winning a point, however, Nadal also showed decreased 'anxiety' (-12, 95% CI -20 to -2%) with winning a point. For Novak Djokovic, winning points was associated with decreased intensities in 'anger' (-16, 95% CI -25 to -2%), 'annoyance' (-16, 95% CI -23 to -3%), and 'focus' (-16, 95% CI -25 to -1%). Roger Federer showed no significant emotional reactions to the outcome of points, in keeping with Federer's 'ice-cool' stereotype.

To assess the predictive value of a player's emotions, we used the same logistic model setup as described above but with the outcome of the next point. The odds ratios from this model would thus indicate the associations between the current observed emotional state and the subsequent point outcome.

For two players, Rafael Nadal and Novak Djokovic, multiple significant associations were found between their predicted emotions and the outcome of the next point (right column, Figure 4). When Nadal was in a state of greater observed 'anger' (-20, 95% CI -29 to -8%), 'annoyance' (-14, 95% CI -21 to -5%), or 'dejection' (-12, 95% CI -20 to -1%), it was associated with a decreased likelihood of winning the next point; whereas a greater state of 'focus' was associated with increased odds of winning the next point (11, 95% CI 1 to 24%). For Novak Djokovic, increases in several emotional states were linked to increased win chances, which included ‘annoyance’ (21, 95% CI 8 to 36%), ‘anxiety’ (15, 95% CI 2 to 30%), ‘dejection’ (13, 95% CI 1 to 27%), and ‘focus’ (15, 95% CI 3 to 30%). Although one significant positive association was found between a state of ‘elation’ and winning the point (17, 95% CI 3 to 34%) for Roger Federer, we found much weaker emotional links with future performance in general for Federer and Andy Murray.

6. Conclusions

This paper presents a framework for predicting emotions in sport for athletes from single-camera video. The proposed framework builds on available image processing tools for facial landmark detection and public data on basic emotional states to develop emotional prediction models tailor-
made for competitive environments. The application of these models to expression data for four of the most well-known among active male tennis players provides the first comprehensive set of quantitative data on the emotional states of athletes in competition. The emotional profiles we observed demonstrated the face validity of the method and we are the first to report a direct link between performance outcomes and the on-court emotions of tennis players.

While the present study has demonstrated the potential of our framework, improving the robustness of the method with further training would be valuable for its future use in applied research and would be a critical step toward introducing emotion detection into real-time sports analytics tools. Some important topics for further model development should include the performance characteristics by player subgroups (e.g. gender, race, etc.) and environmental factors (e.g. light and dark contrast, face angle, etc.). Another direction of research could also look at the comparative performance of methods that use image pixels as their feature set rather than the summary set of facial landmarks derived from facial images. Further, treating the emotion classification problem as a semi- or un-supervised learning problem could help to protect against mislabeling errors and to identify the general emotional features relevant to sport, two advantages that would be worthwhile to explore.

Although this paper has focused on the prediction of emotions in tennis, we believe our framework can be generalized to other sports. The major challenge to extending this work to other sports is the collection of accurately labeled data and the extraction of suitable facial expressions from game broadcasts. Both tasks are currently labor intensive and can be more difficult in sports where play is more continuous, and player faces are frequently obscured or otherwise difficult to detect from standard broadcast. Given these issues, sports in which athlete expressiveness is easy to observe and a common interest during performance would be the most natural starting points for extending our method.

References


