Exploration of Judicial Facial Expression in Videos and Transcripts of Legal Proceedings

A thesis submitted for the degree of

Bachelor of Commerce (Honours)

by

Huize Zhang

27478343

Department of Econometrics and Business Statistics

Monash University

Australia

September 2019

Contents

Chapter 1

Introduction

Decisions by courtroom Justices have been discussed broadly in the legal literature. Gender, political views and religious background of both the Justices and counsel in the case potentially influence the decisions. This paper will explore the facial behaviour of the Justices during hearings with the objective of being to assess whether it can help to predict outcomes. Audio Visual(AV) recordings and case transcripts will be computationally processed and analysed to examine the decisions of each Justice.

1.1 Motivation

Tutton, Mack, and Roach Anleu (2018) attemped to utilize the AV technology, which is made available online by the High Court of Australia (Australia, 2019). They visually inspect of the videos to highlight when Justices depart from the expected norms of judicial conduct. To better understand the emotion status and therefore the departure of the emotional behaviour, more advanced technologies could be applied. An example is to use OpenFace (2018) technology, which provides information on emotions exhibited by the Justices. This technique has been applied by Kovalchik and Reid (2018) on professional tennis players during Grand Slam matches. That study demonstrated the potential to predict the outcome of High Court appeals based on Justices' demeanour utilising contemporary tools and emotion tagging techniques.

1.2 Literature review

The literature sumary is divided into two parts: (1) current work in legal studies to understand the behaviour of the Justices and (2) existing facial recognition and emotion tagging technology.

1.2.1 Legal study from a behaviour perspective

People have attempted to predict the decisions of the Justices for centuries. Chief Justices of Australia and Zealand (2017) present the following code of conduct:

> It is important for judges to maintain a standard of behaviour in court that is consistent with the status of judicial office and does not diminish the confidence of litigants in particular, and the public in general, in the ability, the integrity, the impartiality and the independence of the judge.

This impartiality should be clear in judicial demeanour (Tutton, Mack, and Roach Anleu, 2018; Goffman, 1956). Paul Ekman (Ekman et al., 1991) suggests that from a behavioural perspective, some facial and vocal inflections are often unbeknown to the speakers themselves . Many scholars have exploited this in studying the court outcomes through the language and words used by the Justices in the court (Shullman, 2004) and vocal and facial characteristics of the Justices (Chen et al., 2018).

There are also existing works to understand the emotion of the Justices from a linguistic perspective and suggest some factors that could be useful to indicate how the Justices' vote and thus the court outcome. These factors include the use of pleasant and unpleasant language by Black et al. (2011), the frequency and content of Justices' questions by Shullman (2004) and Johnson et al. (2009). Epstein, Landes, and Posner (2010) use a regression analysis with the number of questions asked by the Justices used to infer the winning party in a case.

Other scholars (Chen, Halberstam, and Alan, 2016; Chen, Halberstam, Yu, et al., 2017; Schubert et al., 1992) have studied the emotion of the Justices from vocal characteristics and suggest that these vocal characteristics, especially perceived masculinity is strongly correlated with the court outcomes. Dietrich, Enos, and Sen (2019) uses a multilevel logistic model with random effects to suggest that subconscious vocal inflections contain information that is not available from text.

Chen (2018) employed both vocal and facial characteristics to predict the court votes using Supreme Court data from 1946-2014. The audio clips are first preprocessed to get the Mel-frequency Cepstral Coefficients (MFCC) and then applied to a random forest model. The image features are extracted using a Histogram of Oriented Gradients (HOG) method. More specific facial recognition software is readily available to extract human facial features and these facial recognition technologies have not yet been applied to the legal proceedings.

Most of the literature is conducted using the U.S. Supreme Court Database and less studies have been conducted using Australian High Court data. Tutton, Mack, and Roach Anleu (2018) have used an ethnographic approach to study the transcript and AV recordings in the High Court of Australia but the study is conducted in an observational manner via matching the Justices' distinct behaviour with the transcript.

1.2.2 Facial recognition

An anatomical analysis of facial action (Ekman and Friesen, 1976) led to the Facial Action Code (FAC) (Ekman and riesen, 1978) and has been further revised by Ekman, Friesen, and Hager (2002). This decomposition of facial muscles is widely used in scientific research. It was applied in competitive sports, specifically tennis by Kovalchik and Reid (2018) who found that the emotion of professional tennis players will have an impact on their performance.

There have been many algorithms created for facial detection and the analysis of their performance when applied to images have been the focus of events like the Audio/Visual Emotion Challenge (Schuller et al., 2012, 2011) and Emotion Recognition In The Wild Challenge and Workshop (Dhall et al., 2013; Kahou et al., 2013).

Facial recognition software has also been implemented by DeepFace (Taigman et al., 2014) from Facebook, and FaceNet (Schroff, Kalenichenko, and Philbin, 2015) from Google.

OpenFace (Baltrusaitis et al., 2018) is the first open-sourced face recognition software that provides facial expression detection, including facial landmarking, head pose estimation, eye gaze tracking and facial action unit detection. Along with its previous version (Baltrušaitis, Robinson, and Morency (2016)), the OpenFace toolkit has been used in different social research studies including depression classification (Yang et al. (2016) and Nasir et al. (2016)).

Chapter 2

Data Collection

2.1 Data Processing

The source data for this research project is the AV recordings publicly available from the High Court of Australia (Australia, 2019). Multiple procedures need to be performed to obtain the dataset.

The workflow for extracting numerical data from the videos can be found in Figure 2.1. Youtube-dl (Hsuan, Amine, and Sergey, 2019) has been used to download videos from the High Court of Australia(Australia, 2019) and the list of videos used in this research project is documented in the Appendix. Image frames are extracted from the videos for every minute via ffmpeg (Bellard, 2019), resulting in 1021 image frames (252 frames from `Nauru` videos and 769 frames from other five videos). Taipan (Kobakian and O'Hara-Wild, 2018) is then used to find the x-y coordinates of the location of the Justices in each image frame. ImageMagick (Cristy et al., 2019) is then used to crop the face of each Justice from each image frame that is taken from each video where three Justices present in `Nauru` videos and five Justices in other videos. The resulting 4601 cropped images (252 image frames from `Nauru` videos where three Justices are presented and 769 image frames from other videos where five justices are presented) are then sent to OpenFace (Baltrusaitis et al., 2018) to produce the variables for facial landmarking, head pose, eye gaze and facial action unit. This step is performed via the docker platform. The resulting outputs from

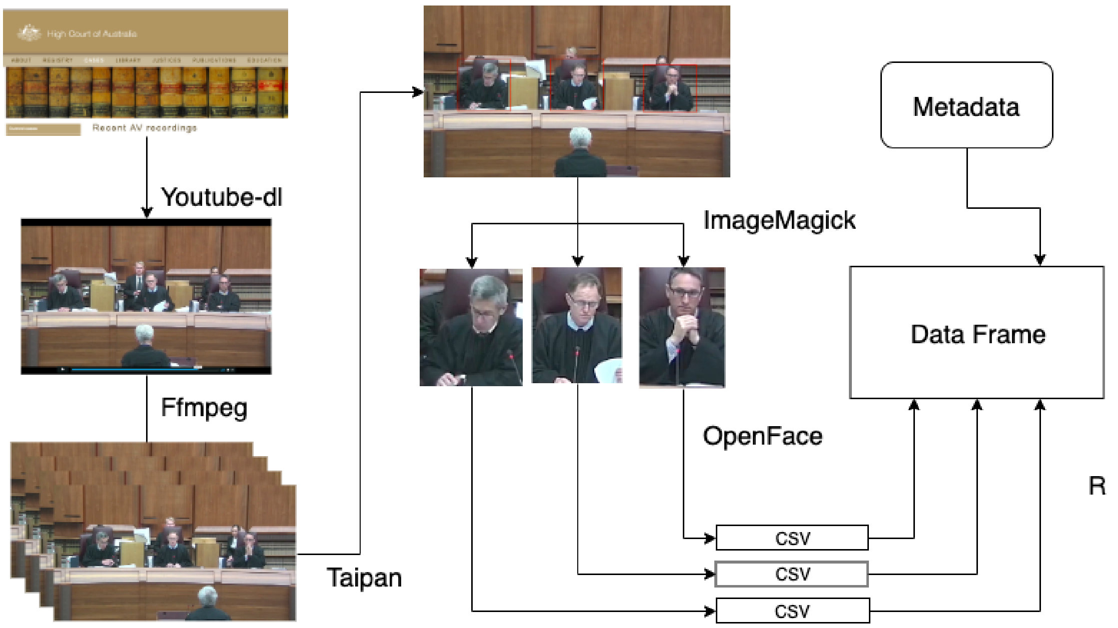


Figure 2.1: *workflow*

OpenFace are individual comma-separated values (csv) files for each of the 4601 faces considered and processing is done in R to combine all the separate csv files into a final dataframe with appropriate index of `frame_id`, `judge_id` and `video_id`.

2.2 Variable description

OpenFace provides more than 711 variables measuring different aspect of a given face and a full description of the output variables can be found here. This outlines the difficulty of this project: no existing models will present accurate prediction and inference using 700+ variables - how can we incorporate these information to say about the facial expressions of the Justices during the hearings?

I conduct some exploratory data analysis on one video: `Nauru_a` and find the 700+ variables can be classified as follows with some insights

- **Confidence**: How confidence OpenFace is with the detection. Confidence is related to the angle that the Justice's face present in the images.
- **Gaze**: Gaze tracking: the vector from the pupil to corneal reflection. The dataset contains information on the gaze for both eyes while there is no distinct difference

between the eyes. Also I was trying to make animation to track the change of the gaze for judges but no good luck.

- **Pose**: the location of the head with respect to camera. Pose-related variables don't provide much useful information apart from gaze-related variables.

- **Landmarking**: landmarking variables for face and eyes. Landmarking variables allows me to plot the face of the judge in a particular frame. More work could be done to explore the usefulness of landmarking variables.

- **Action Unit**: Action units are used to describe facial expressions. this website provides a good animation on each action unit. The action unit has intensity measures ending with `_c` and presence measures ending with `_r`. These variables will be the focus of my project and a reference study of using action units to detect human emotion by Kovalchik can be found here.

2.3 Missing value imputation

The missingness in the dataset could be due to the fact that a judge is reading the materials on the desk so the face is not captured for a particular frame or simply because some faces are not detectable for the given resolution of the video stream. However, since that data is in time series structure, simply drop the missing observation will cause the time interval to be irregular and complicate further analysis.

There are two different sets of variables that need imputation. `Presence` is a binary variable that takes value of one if an action unit is present in a particular frame for a judge in a video and `Intensity` measures how strong that action unit is. Linear interpolation from `forecast` package is suitable to impute `Intensity` and `Presence` is imputed through sampling from binomial distribution. The imputed action unit data is stored as `au_imputed` under the `raw_data` folder.

2.4 Data quality

There is a data quality issue coming from the data I get from OpenFace. For some observations, the intensity of the action unit could be high while the present variable has a zero value. This does not make sense since if an action unit has been detected as strong intensity for a judge in a particular frame, it should at least present on the judge's face. Therefore, I adjust for the presence value if the intensity is higher than one. One is being chosen as the threshold value since in Ekman's definition of the intensity of the action unit, a score of one means the action unit is at least slightly present in the judge's face. The adjusted data is stored as `au_tidy` under the `raw_data` folder.

Chapter 3

Methods

Here you need to write about the analytical methods that you are using

Chapter 4

Results

4.1 Notation

Let X_{jitkl} be a matrix of predictors, and Y_{jitkl} variable in our case is bivariate matrix of response variables, including a binary indicator of presence/absence and a numeric value measuring intensity, of facial action unit, where

- i indicates `judge_id` and $i = 1, 2, \cdots, 6$
- j indicates `video_id` and $j = 1, 2, \cdots, 7$
- t indicates `frame_id` and $t = 1, 2, \cdots, T_j$
- k indicates `au_id` and available action unit includes AU01, AU02, AU04, AU05, AU06, AU07, AU09, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, AU28, AU45. Notice that OpenFace doesnt provide intensity score for AU28.
- l indicates appellant or respondent, $l = 1, 2$

Note that t could be considered a time variable, but because images are taken at 1 minute intervals, temporal dependence is unlikely to exist. Rather this should be considered an independent observation.

Assuming all the facial information can be summarised as a `Y` variable with multiple indices (i, j, t, k, l). A main effect model for this data can be written as

$$\widehat{Y}_{ijtk} = \mu + \alpha_i + \beta_j + \gamma_t + \delta_k + \psi_l$$

Also, let P_{jitkl} represent the response variable presence, and I_{jitkl} represent the second response variable intensity.

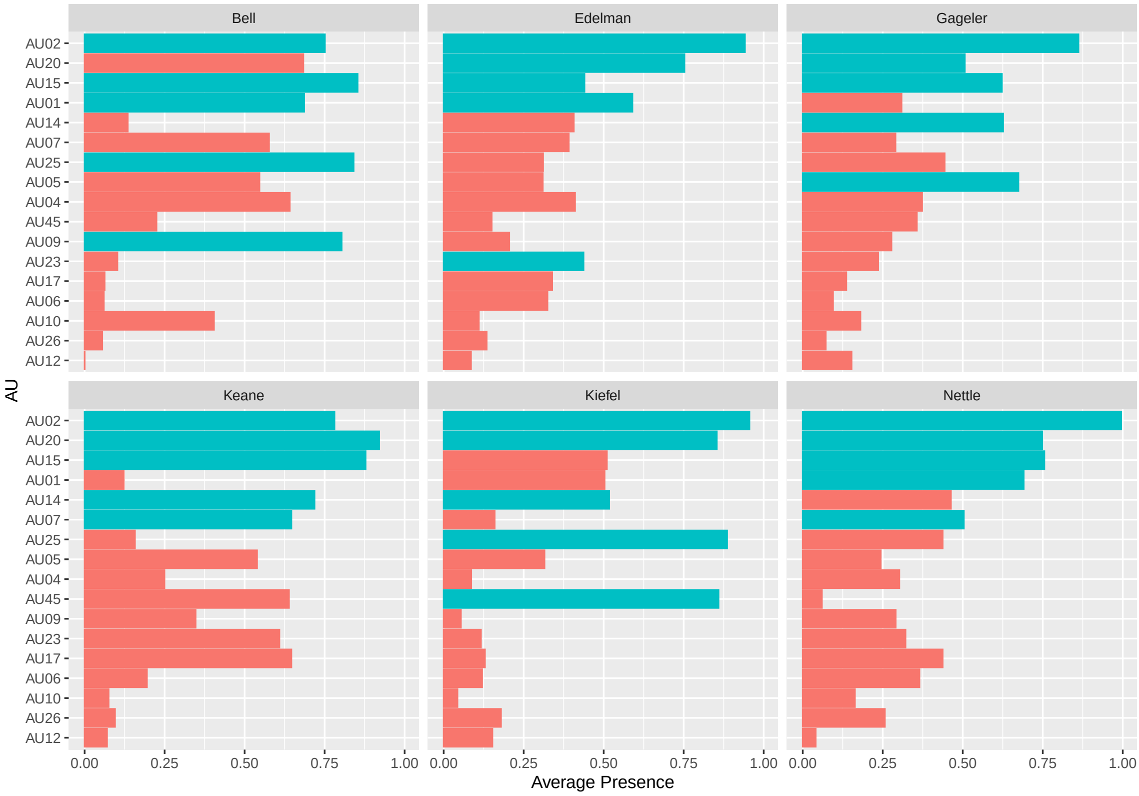
4.2 Action unit: Presence

4.2.1 Mean presence

The plot gives an overview of the presence of all the action unit across all the judge. The statistic each bar represents is the average presence of an action unit for a judge throughout all the video time and it can be written as

$$P_{ik} = \frac{\sum_{jt} X_{ijtk}}{\sum_{j=1}^{J} T_j}$$

. The order of Action unit on the y axis is ranked by the average presence of all the judge, which can be re-presented as P_{*K}.



The most frequent displayed action unit is highlighed in blue for each judge and summarised in the table below.

index	Bell	Edelman	Gageler	Keane	Kiefel	Nettle
1	AU15	AU02	AU02	AU20	AU02	AU02
2	AU25	AU20	AU05	AU15	AU25	AU15
3	AU09	AU01	AU14	AU02	AU45	AU20
4	AU02	AU15	AU15	AU14	AU20	AU01
5	AU01	AU23	AU20	AU07	AU14	AU07

It can be seen that some of the action units are common across almost all the judges, these includes

- AU02 (outer eyebrow raise),
- AU20 (lip stretcher),
- AU15 (Lip Corner Depressor) and
- AU14 (Dimpler)

The table below summarises the judge specific high frequent action units.

judge_id	first	second	3
Bell	AU25	AU09	AU01
Edelman	AU01	AU23	NA
Gageler	AU05	NA	NA
Keane	AU07	NA	NA
Kiefel	AU25	AU45	NA
Nettle	AU01	AU07	NA

These are the results from inspecting the action units visually and they should also be reflected thorugh the coefficients of the relevant models [see stage 3 modelling].

4.2.2 Model fit

The first model I use is a generalised linear model with binomial link to understand the presence of the action units. The variables used include the judge_id and a selection of action units based on the mean presence result in the previous section. The action units are chosen to be the ones with at least two judges have them to be the most common five action units. The model can be written down as

$$P_{ik} = \mu + \alpha_i + \delta_k + (\alpha\delta)_{ik}$$

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual
35945.66	26571	-14756.03	29596.06	29939.94	29512.06	26530

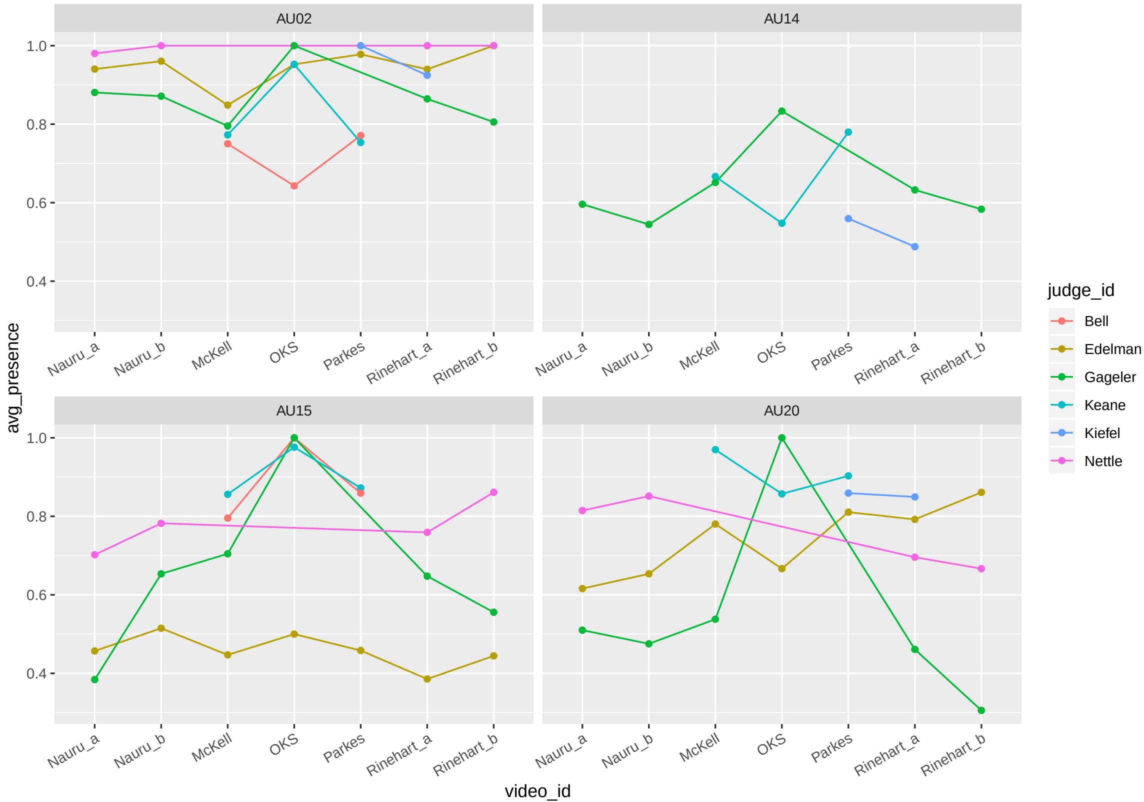
term	estimate	std.error	statistic	p.value
(Intercept)	-0.80	0.07	-11.85	0.00
judge_idBell	0.99	0.12	8.16	0.00
judge_idGageler	1.53	0.10	15.05	0.00
judge_idKeane	0.96	0.12	7.92	0.00
judge_idKiefel	0.02	0.11	0.22	0.83
judge_idNettle	-0.33	0.12	-2.86	0.00
AUAU01	1.16	0.09	12.54	0.00
AUAU02	3.58	0.15	23.95	0.00
AUAU14	0.42	0.09	4.56	0.00
AUAU15	0.56	0.09	6.06	0.00
AUAU20	1.91	0.10	19.29	0.00
AUAU25	0.00	0.10	0.05	0.96
judge_idBell:AUAU01	-0.57	0.17	-3.27	0.00
judge_idGageler:AUAU01	-2.70	0.14	-18.95	0.00
judge_idKeane:AUAU01	-3.29	0.20	-16.08	0.00
judge_idKiefel:AUAU01	-0.38	0.16	-2.43	0.02
judge_idNettle:AUAU01	0.77	0.16	4.88	0.00
judge_idBell:AUAU02	-2.66	0.21	-12.44	0.00
judge_idGageler:AUAU02	-2.47	0.20	-12.60	0.00
judge_idKeane:AUAU02	-2.46	0.22	-11.37	0.00
judge_idKiefel:AUAU02	0.26	0.27	0.98	0.33
judge_idNettle:AUAU02	2.88	0.60	4.77	0.00
judge_idBell:AUAU14	-2.47	0.20	-12.33	0.00
judge_idGageler:AUAU14	-0.63	0.14	-4.51	0.00
judge_idKeane:AUAU14	0.36	0.18	2.03	0.04
judge_idKiefel:AUAU14	0.42	0.16	2.72	0.01
judge_idNettle:AUAU14	0.56	0.15	3.63	0.00
judge_idBell:AUAU15	1.01	0.20	5.16	0.00
judge_idGageler:AUAU15	-0.79	0.14	-5.61	0.00
judge_idKeane:AUAU15	1.26	0.20	6.14	0.00
judge_idKiefel:AUAU15	0.26	0.15	1.66	0.10
judge_idNettle:AUAU15	1.70	0.16	10.52	0.00
judge_idBell:AUAU20	-1.33	0.18	-7.50	0.00

4.2.3 Presence by videos

Apart from visualising the general presence score for all the action unit, we are also interested in the break down statistics by video. The statistics being plotted is thus

$$P_{ijk} = \frac{\sum_t X_{ijtk}}{T_j}$$

with selected most common four action units. From this plot, it is interesting to know that almost all the judges have more frequent action units on the face for case OKS. This maybe related to the nature of the case...



4.2.4 Model fit

$$P_{ijk} = \mu + \alpha_i + \beta_j + \delta_k + (\alpha\beta)_{ij} + (\alpha\delta)_{ik} + (\beta\delta)_{jk}$$

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual
35945.66	26571	-14525.62	29245.24	30039.44	29051.24	26475

term	estimate	std.error	statistic	p.value
(Intercept)	-0.72	0.13	-5.55	0.00
judge_idBell	0.80	0.15	5.30	0.00
judge_idGageler	1.31	0.14	9.65	0.00
judge_idKeane	1.19	0.15	7.87	0.00
judge_idKiefel	-0.02	0.13	-0.14	0.89
judge_idNettle	-0.66	0.15	-4.36	0.00
video_idNauru_b	-0.14	0.19	-0.73	0.46
video_idMcKell	-0.03	0.17	-0.17	0.87
video_idOKS	-2.20	0.29	-7.67	0.00
video_idParkes	-0.13	0.16	-0.79	0.43
video_idRinehart_a	0.06	0.14	0.46	0.65
video_idRinehart_b	-0.17	0.26	-0.64	0.52
AUAU01	1.53	0.17	9.03	0.00
AUAU02	3.66	0.26	13.95	0.00
AUAU14	0.51	0.16	3.10	0.00
AUAU15	0.17	0.17	1.03	0.30
AUAU20	1.88	0.17	10.87	0.00
AUAU25	0.41	0.17	2.49	0.01
judge_idGageler:video_idNauru_b	-0.04	0.15	-0.25	0.81
judge_idNettle:video_idNauru_b	0.12	0.16	0.73	0.47
judge_idBell:video_idMcKell	0.58	0.14	4.17	0.00
judge_idGageler:video_idMcKell	0.41	0.14	2.87	0.00
judge_idKeane:video_idMcKell	-0.24	0.15	-1.61	0.11
judge_idBell:video_idOKS	2.26	0.27	8.40	0.00
judge_idGageler:video_idOKS	1.29	0.23	5.60	0.00
judge_idKeane:video_idOKS	0.65	0.25	2.62	0.01
judge_idKiefel:video_idParkes	-0.07	0.11	-0.65	0.52
judge_idGageler:video_idRinehart_a	0.35	0.12	3.01	0.00
judge_idNettle:video_idRinehart_a	0.26	0.12	2.14	0.03
judge_idGageler:video_idRinehart_b	-0.16	0.22	-0.76	0.45
judge_idNettle:video_idRinehart_b	1.23	0.24	5.07	0.00
judge_idBell:AUAU01	-0.46	0.20	-2.35	0.02
judge_idGageler:AUAU01	-2.78	0.15	-18.70	0.00

4.2.5 Appellent vs. Respondent

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual
35945.66	26571	-14476.17	29244.35	30439.74	28952.35	26426

term	estimate	std.error	statistic	p.value
(Intercept)	-0.69	0.16	-4.42	0.00
judge_idBell	0.80	0.15	5.32	0.00
judge_idGageler	1.31	0.14	9.62	0.00
judge_idKeane	1.20	0.15	7.89	0.00
judge_idKiefel	-0.02	0.13	-0.14	0.89
judge_idNettle	-0.66	0.15	-4.35	0.00
AUAU01	1.49	0.21	7.18	0.00
AUAU02	3.28	0.30	11.10	0.00
AUAU14	0.47	0.20	2.33	0.02
AUAU15	0.02	0.20	0.11	0.91
AUAU20	1.55	0.21	7.42	0.00
AUAU25	0.47	0.20	2.34	0.02
video_idNauru_b	-0.19	0.21	-0.88	0.38
video_idMcKell	0.05	0.21	0.22	0.82
video_idOKS	-1.81	0.34	-5.36	0.00
video_idParkes	-0.37	0.21	-1.77	0.08
video_idRinehart_a	-0.10	0.18	-0.53	0.60
video_idRinehart_b	-0.07	0.31	-0.22	0.83
speakerRespondent	-0.08	0.21	-0.35	0.72
judge_idBell:AUAU01	-0.47	0.20	-2.38	0.02
judge_idGageler:AUAU01	-2.78	0.15	-18.72	0.00
judge_idKeane:AUAU01	-3.37	0.23	-14.88	0.00
judge_idKiefel:AUAU01	-0.36	0.16	-2.20	0.03
judge_idNettle:AUAU01	0.71	0.17	4.23	0.00
judge_idBell:AUAU02	-2.93	0.25	-11.60	0.00
judge_idGageler:AUAU02	-2.47	0.20	-12.16	0.00
judge_idKeane:AUAU02	-2.71	0.25	-10.71	0.00
judge_idKiefel:AUAU02	0.31	0.28	1.12	0.26
judge_idNettle:AUAU02	3.02	0.61	4.94	0.00
judge_idBell:AUAU14	-2.83	0.23	-12.51	0.00
judge_idGageler:AUAU14	-0.70	0.15	-4.80	0.00
judge_idKeane:AUAU14	0.19	0.20	0.94	0.35
judge_idKiefel:AUAU14	0.59	0.16	3.62	0.00
judge_idNettle:AUAU14	0.64	0.16	[illegible]	[illegible]

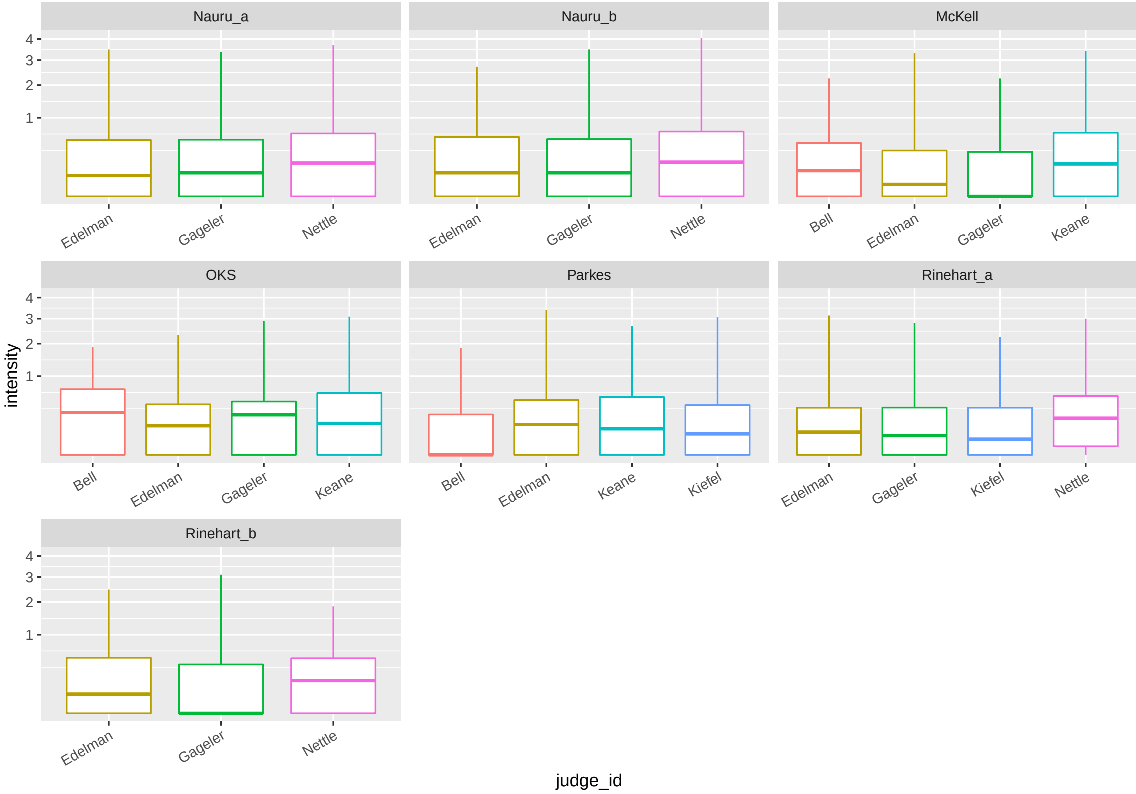
4.3 Action unit: Intensity

4.3.1 General Intensity plot

The plot gives an overview of the action unit intensity of all the judges across all the trails. Each bar-and-whisker represents the intensity of all the action units aggregated on time for a particular judge in a specific case. For example, the first bar-and-whisker in case Nauru_a is created using all the 17 action units of Edelman through out the elapsed time in Nauru_a case. In mathematics notation, the plotted statistics is I_{ijtk} seperating by $i$and$j$.

In Ekman's 20002 FACS manual, the intensity of Action unit is defined based on five classes: Trace(1), Slight(2), Marked or pronounced(3), Severe or extreme(4) and Maximum(5). From the plot, most of the action units have low intensity (almost zero average and lower than one upper bounds) and this is expected because usually in the court room, judges are expected to behave neutral. From this plot, we can see that Judge Bell doesn't seem to have many intensive expressions as we can see from the relatively small amount of dots in the whisker.

To better look at the mean of each boxplot, we take a square root transformation and hide the outliners into the upper line. We can find that Judge Nettle seems to have higher average in all the four cases he appears: Nauru_a&b, Rinehart_a &b.

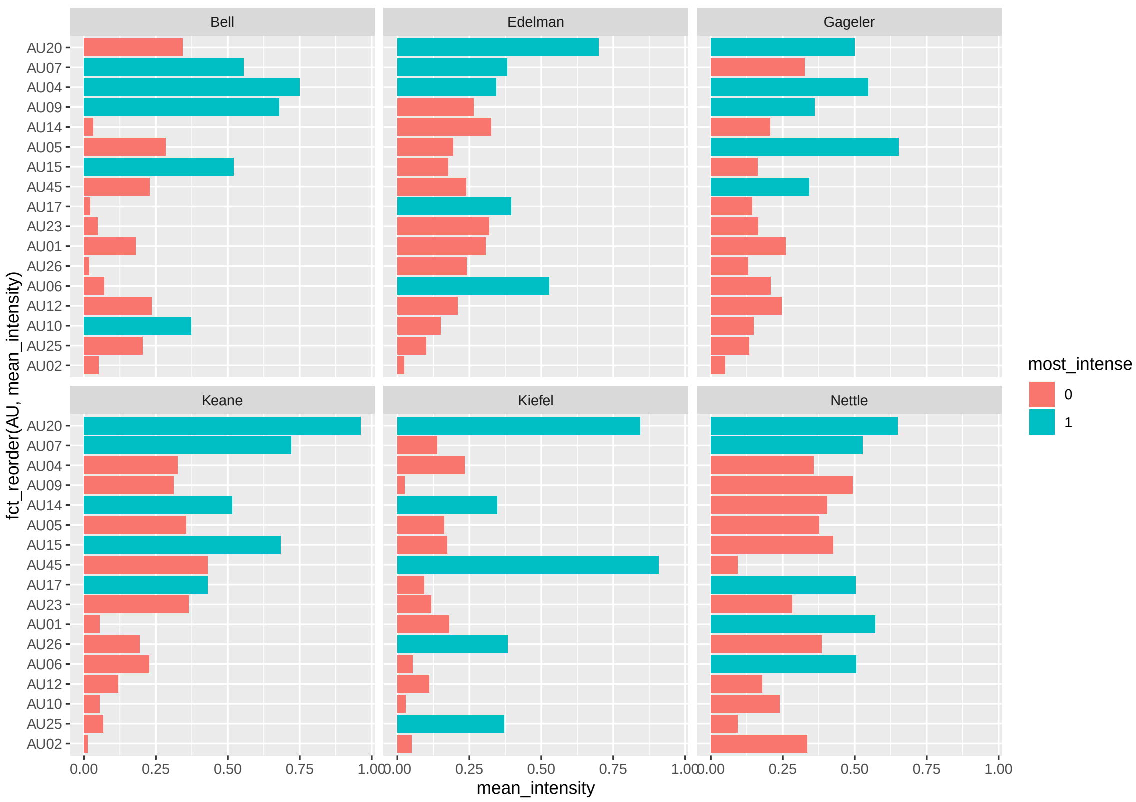


4.3.2 Mean intensity

We compute a similar mean intensity score for each of the action unit for each of the judge, the statsitics is

$$I_{ik} = \frac{\sum_{jt} X_{ijtk}}{\sum_{j=1}^{J} T_j}$$

. Less uniform as the mean presence plot - different judge response intense at different action units.

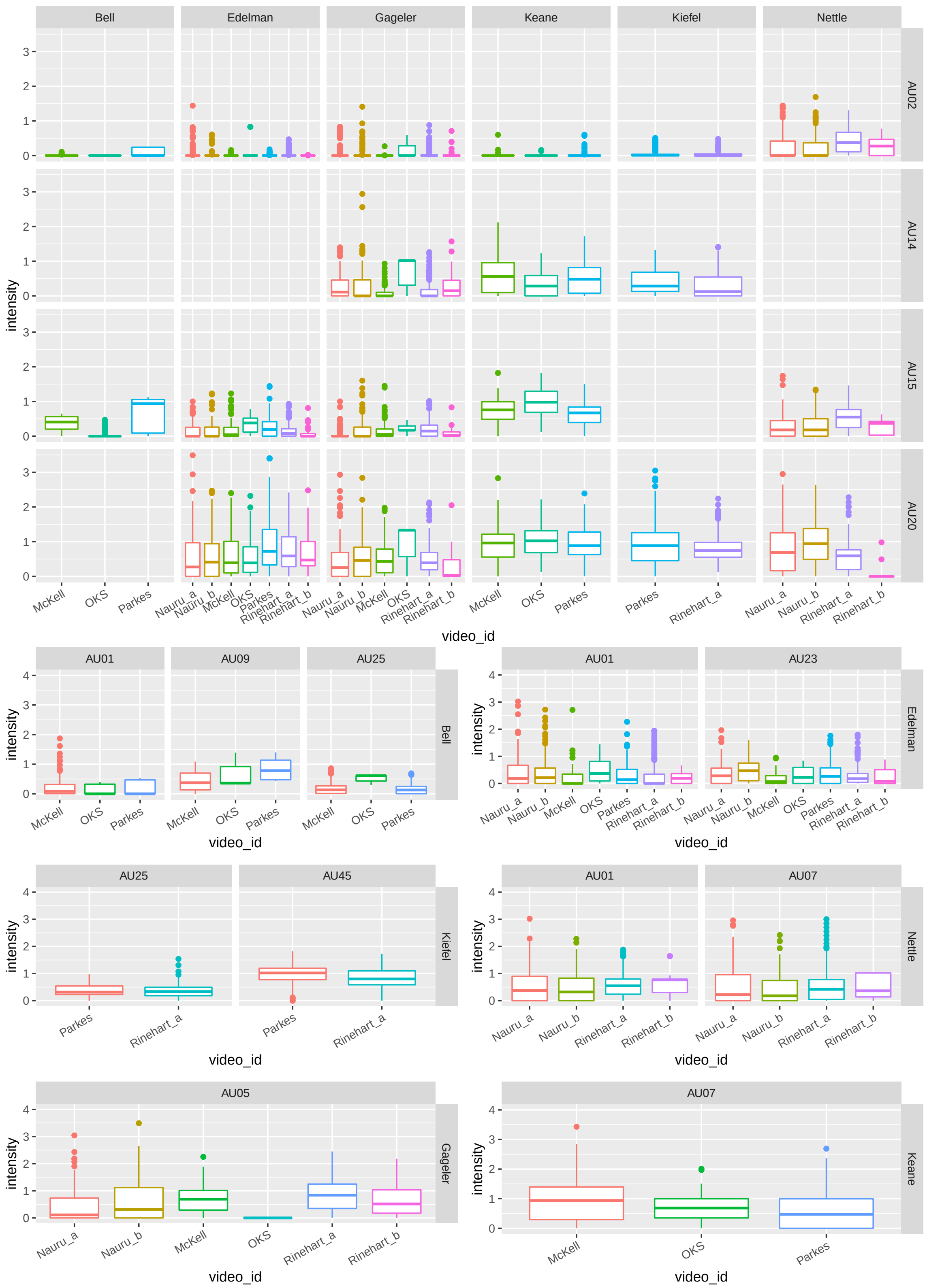


4.3.3 Intensity plot for the most frequent action units

Apart from visualising the general intensity score for all the action unit, we are also interested in the intensity score of the most frequent units. The statistics being plotted is thus I_{ij*k} with selected i indicated by the `most_common` table above. Five most frequent action units for each of the six judges give 30 subplots and I arrange them into two seperate plots: Intensity for four major high frequent action units and Intensity for other high frequent action units.

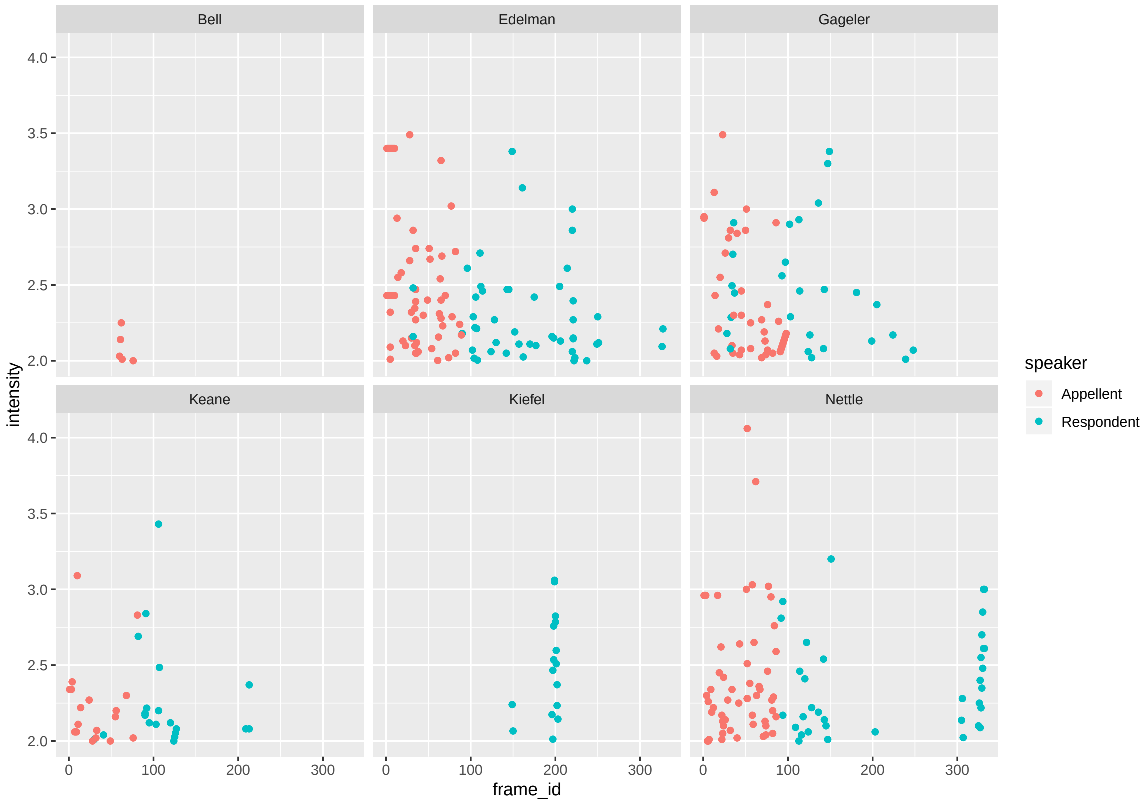
From the first plot, some panels are empty for some judges. This is because the particular action unit is not the most frequent five for that particular judge. We can learn that AU02, although being commonly detected for all the judges, the intensity is quite low.

From the second plot, most of the intensity are low as the general intensity plot, while one action unit has drawn the attention: AU45 for Kiefel. The intensity in both case parkes and case Rinehart_a for Kiefel are relatively higher than others and this will be reflected through our model.


Bell
Edelman
Gageler
Keane
Kiefel
Nettle
AU02
AU14
AU15
AU20
intensity
McKell
OKS
Parkes
Nauru_a
Nauru_b
Rinehart_a
Rinehart_b
video_id
AU01
AU09
AU25
AU23
AU45
AU07
AU05

4.3.4 High intensity points

We filter out the points have intensity greater than 2 (at least "slight" as per Ekman) in the previous plot and plot it against time and color by the speaker. It tells us that Edelman, Gageler and Nettle are the judges have stronger emotion that can be detected (since they have more points with intensity greater than 2). Different judges also have different time where they display stronger emotions. For example, Justice Nettle are more likely to have stronger emotion throughout the time when the appellant is speaking but only at the beginning and ending period when the respondent is speaking.



Appendix A

Additional stuff

Table A.1: *Details of videos processed.*

Case	Name	AV recording link
Republic of Nauru v. WET040	Nauru_a	http://www.hcourt.gov.au/cases/cases-av/av-2018-11-07a
TTY167 v. Republic of Nauru	Nauru_b	http://www.hcourt.gov.au/cases/cases-av/av-2018-11-07b
Rinehart & Anor v. Hancock Prospecting Pty Ltd & Ors on 13 Nov 18	Rinehart_a	http://www.hcourt.gov.au/cases/cases-av/av-2018-11-13
Rinehart & Anor v. Hancock Prospecting Pty Ltd & Ors on 14 Nov 18	Rinehart_b	http://www.hcourt.gov.au/cases/cases-av/av-2018-11-14a
Parkes Shire Council v. South West Helicopters Pty Limited	Parkes	http://www.hcourt.gov.au/cases/cases-av/av-2018-11-14b

Case	Name	AV recording link
McKell v. The Queen	`McKell`	http://www.hcourt.gov.au/cases/cases-av/av-2018-12-07
OKS v. The State of Western Australia	`OKS`	http://www.hcourt.gov.au/cases/cases-av/av-2019-02-14

Bibliography

Australia, HC of (2019). *Recent AV recordings*. Accessed: 2019-05-03. http://www.hcourt.gov.au/cases/recent-av-recordings.

Baltrušaitis, T, P Robinson, and LP Morency (2016). Openface: an open source facial behavior analysis toolkit. In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, pp.1–10.

Baltrusaitis, T, A Zadeh, YC Lim, and LP Morency (2018). Openface 2.0: Facial behavior analysis toolkit. In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, pp.59–66.

Bellard, F (2019). *ffmpeg*. https://ffmpeg.org/.

Black, RC, SA Treul, TR Johnson, and J Goldman (2011). Emotions, oral arguments, and Supreme Court decision making. *The Journal of Politics* **73**(2), 572–581.

Chen, DL, M Kumar, V Motwani, and P Yeres (2018). *Is Justice Really Blind? And Is It Also Deaf*. Tech. rep. Technical report.

Chen, D, Y Halberstam, and C Alan (2016). Perceived masculinity predicts us supreme court outcomes. *PloS one* **11**(10), e0164324.

Chen, D, Y Halberstam, A Yu, et al. (2017). Covering: Mutable characteristics and perceptions of voice in the US Supreme Court. *Review of Economic Studies invited to resubmit, TSE Working Paper* (16-680).

Chief Justices of Australia, TC of and N Zealand (2017). *Guide to Judicial Conduct*. 3rd. Melbourne: Austral-asian Institute of Judicial Administration.

Cristy, J, D Lemstra, G Randers-Pehrson, and B Roucries (2019). *ImageMagick-dl*. https://github.com/ImageMagick.

Dhall, A, R Goecke, J Joshi, M Wagner, and T Gedeon (2013). Emotion recognition in the wild challenge 2013. In: *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, pp.509–516.

Dietrich, BJ, RD Enos, and M Sen (2019). Emotional arousal predicts voting on the US supreme court. *Political Analysis* **27**(2), 237–243.

Ekman, P and WV Friesen (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior* **1**(1), 56–75.

Ekman, P, WV Friesen, and JC Hager (2002). Facial action coding system: The manual on CD ROM. *A Human Face, Salt Lake City*, 77–254.

Ekman, P, M O'Sullivan, WV Friesen, and KR Scherer (1991). Invited article: Face, voice, and body in detecting deceit. *Journal of nonverbal behavior* **15**(2), 125–135.

Ekman, P and WV riesen (1978). *Facial action coding system*. Palo Alto: CA: Consulting Psychologists Press.

Epstein, L, WM Landes, and RA Posner (2010). Inferring the winning party in the Supreme Court from the pattern of questioning at oral argument. *The Journal of Legal Studies* **39**(2), 433–467.

Goffman, E (1956). The nature of deference and demeanor. *American Anthropologist* **58**(3), 473–502.

Hsuan, YC, R Amine, and M Sergey (2019). *youtube-dl*. https://github.com/ytdl-org/youtube-dl/.

Johnson, TR, RC Black, J Goldman, and SA Treul (2009). Inquiring minds want to know: Do justices tip their hands with questions at oral argument in the US supreme court. *Washington University Journal of Law and Policy* **29**, 241.

Kahou, SE, C Pal, X Bouthillier, P Froumenty, Ç Gülçehre, R Memisevic, P Vincent, A Courville, Y Bengio, RC Ferrari, et al. (2013). Combining modality specific deep neural networks for emotion recognition in video. In: *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, pp.543–550.

Kobakian, S and M O'Hara-Wild (2018). *taipan: Tool for Annotating Images in Preparation for Analysis*. R package version 0.1.2. https://CRAN.R-project.org/package=taipan.

Kovalchik, S and M Reid (2018). Going inside the inner game: Predicting the emotions of professional tennis players from match broadcasts. In: MIT Sloan Sports Analytics Conference. Accessible at http://www ...

Nasir, M, A Jati, PG Shivakumar, S Nallan Chakravarthula, and P Georgiou (2016). Multimodal and multiresolution depression detection from speech and facial landmark features. In: *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. ACM, pp.43–50.

Schroff, F, D Kalenichenko, and J Philbin (2015). Facenet: A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.815–823.

Schubert, JN, SA Peterson, G Schubert, and S Wasby (1992). Observing Supreme Court oral argument: A biosocial approach. *Politics and the Life Sciences* **11**(1), 35–52.

Schuller, B, M Valstar, F Eyben, G McKeown, R Cowie, and M Pantic (2011). Avec 2011–the first international audio/visual emotion challenge. In: *International Conference on Affective Computing and Intelligent Interaction*. Springer, pp.415–424.

Schuller, B, M Valster, F Eyben, R Cowie, and M Pantic (2012). AVEC 2012: the continuous audio/visual emotion challenge. In: *Proceedings of the 14th ACM international conference on Multimodal interaction*. ACM, pp.449–456.

Shullman, SL (2004). The illusion of devil's advocacy: How the justices of the supreme court foreshadow their decisions during oral argument. *Journal of Appellate Practice and Process 2* **271**(6).

Taigman, Y, M Yang, M Ranzato, and L Wolf (2014). Deepface: Closing the gap to human-level performance in face verification. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.1701–1708.

Tutton, J, K Mack, and S Roach Anleu (2018). Judicial Demeanor: Oral Argument in the High Court of Australia. *Justice System Journal* **39**(3), 273–299.

Yang, L, D Jiang, L He, E Pei, MC Oveneke, and H Sahli (2016). Decision tree based depression classification from audio video and language information. In: *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. ACM, pp.89–96.