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

October 2019

Contents

Acknowledgements

Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or equivalent institution, and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Huize Zhang

Abstract

Chapter 1

Introduction

1.1 Background

(placeholder)

1.2 Statement of topic

(may go to abstract) 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.

1.3 Motivation

People have attempted to predict the decisions of the Justices in the past century using judge characteristics i.e. Gender, political views, religious background. More recently, scholars(Shullman, 2004; Chen et al., 2018) have been using more information from media(i.e. AV recording, transcript, language used by the Justices) to predict the case outcome using the U.S. Supreme Court data. On-court information has also been used to study data from High Court of Australia. Tutton, Mack, and Roach Anleu (2018) has used an ethnographic approach to present a observational study of judicial behaviour based on watching the audio footage. Manually observing the AV recordings could lead to subjective evaluation of facial expression and this motivates us to build upon Tutton, Mack, and Roach Anleu (2018)'s work to employ facial recognition technology to study the facial expression of the justices, which will provide a more objective result than Tutton, Mack, and Roach Anleu (2018).

1.4 Literature review

The literature summary 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.

1.4.1 Legal study from a behaviour perspective

There is a large law & economics and political science literature that attempts to predict how judges will vote in court cases. Much of this focuses on the characteristics of the judge i.e. gender, political views, religious background and characteristics of the parties in the case i.e. gender or race of the defendant in criminal cases (Nagel, 1962; Koppen and Kate, 1984; Aliotta, 1987-1988; Welch, Combs, and Gruhl, 1988; Steffensmeier and Britt, 2001; Kulik and Perry, 2003).

Moving from static information of the judge and parties involved, more studies start to incorporate the language used by the judge on the court to predict the decision of the Justices. Black et al. (2011) has study the use of pleasant and unpleasant language by the Justices and Shullman (2004) and Johnson et al. (2009) have studied the effect of frequency and content of Justices' questions. 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.

More recent legal study has focused on the usage of emotion and vocal characteristics of the Justices to predict the judge's vots. Although 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.

and this impartiality has been highlighted in judicial demeanour by Tutton, Mack, and Roach Anleu (2018) and 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. Chen, Halberstam, and Alan (2016); Chen, Halberstam, Yu, et al. (2017) and 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) has used a multilevel logistic model with random effects to suggest that subconscious vocal inflections contain information that is not available from text.

Moreover, a more sizeable study by Chen et al. (2018) have incorporated both vocal and image information of the judge into a machine learning model to predict the judge votes and case outcome using the U.S. Supreme Court data from 1946-2014. He found that image features improved prediction of case outcomes from 64% to 69% and audio features improved prediction of case outcomes from 67% to 69%. This demonstrates the potential of incorporating facial information to understand the decision of the Justices.

The literature mentioned above is mostly 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) has used an ethnographic approach to study the judicial demeanour in the High Court of Australia and it is the first of its kind to use transcript and AV recordings in Australian study. The study found that Justices present a detached facial demeanour during the court in most of the time while some human display of emotions i.e. laughter and humour have also been captured by the scholars. Tutton's work has confirmed the potential of using image information to understanding the Justices as in Chen's study, while the ethnographic approach could be biased and lead to subjective results when different people are observing the videos. Thus, building upon Tutton's study, my work fills the gap of producing objective result via utilising facial recognition technology.

1.4.2 Facial recognition

An anatomical study of the decomposition of facial muscles by (Ekman and Friesen, 1976) led to the devlopment of Facial Action Code (FAC) (Ekman and riesen, 1978) and identification of the six universal emotions on human faces. This work has been further revised as (*Facial Action Coding System* n.d.) and has laid a solid foundation for analysing facial expression and developing facial recognition softwares for researchers (Kobayashi and Hara, 1992; Huang and Huang, 1997; Lien et al., 2000; Kapoor, Qi, and Picard, 2003; Tong, Liao, and Ji, 2007; Cohn et al., 2009; Lucey et al., 2010).

To be able to analysis the facial expression, proper facial recognition technique is needed to first extract faces from images. Facial recognition softwares i.e. DeepFace (Taigman et al., 2014) from Facebook and FaceNet (Schroff, Kalenichenko, and Philbin, 2015) from Google have also been developed for face detection. 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. The OpenFace toolkit has been used in different area in research including depression classification (Yang et al., 2016; Nasir et al., 2016b), emotion study (Pan and Hamilton, 2018; Nasir et al., 2016a; Huber et al., 2018) and even sports analytics (Kovalchik and Reid, 2018).

1.5 Research Question

(placeholder)

- Extract facial expression data of the Justices from videos of High Court of Australia
- Merge with data from text transcript
- Statistically model judges facial expressions
- Provide an **objective** source of data to study the problem

- Do the results agree or disagree with Tutton's findings, that the justices are appearing impartial?

Two questions can be explained by the model:

- Do the justices' expression differ from case to case?
- Do the justices' expression differ when different parties are speaking?

The aim of this study is to use facial recognition technology to detect judicial thinking and thus their decisions. There are four specific objectives: • Read in video streams and convert into a numerical data format. • Perform data quality checks to investigate video and data quality. • Exploratory data analysis of the facial expression, trascript variables and outcomes. • Explore models to predict the appeal outcomes based on facial expression and text analysis.

1.6 Significance

(placeholder)

Facial recognition analysis of the videos provides a way to **objectively** and **automatically** assess judicial behaviour.

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). Due to the requirement of resolution (more than 30px for face detection) of OpenFace, we picked up seven cases from 2018 that have less than seven judges as the sample videos for our dataset. A full list of video being processed can be found in Table A.1 in the Appendix.

Multiple procedures need to be performed to obtain the numerical value of facial variables from the source videos. The entire workflow has been plotted 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). 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 followed 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 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

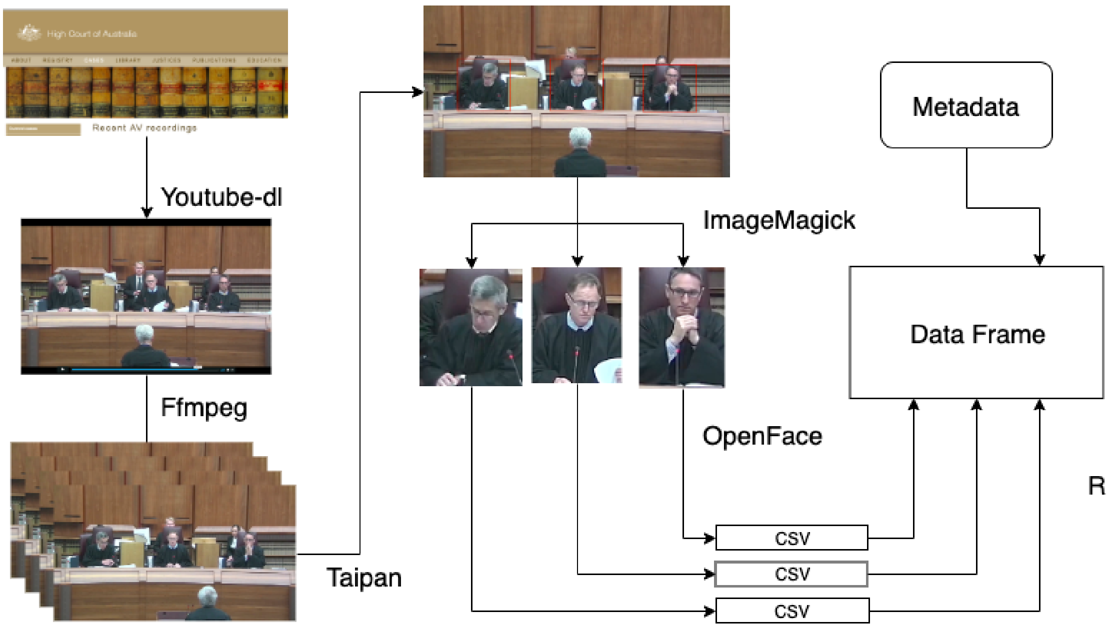


Figure 2.1: *workflow for video and image processing*

resulting outputs from 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`, `judg` and `video`.

2.2 Variable description

OpenFace provides more than 711 variables measuring different aspects of a given face and a full description of the output variables can be found in Baltrusaitis et al. (2018). 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. The action unit has intensity measures ending with `_c` and presence measures ending with `_r`.

2.3 Data format

In this project, we will make use of the action unit variables along with all the added indexes to analyse the face of the judge. In the wide format, apart from the first four index columns, each action unit has two columns with one for binary presence value and another for numeric intensity value. The Table 2.1 presents the first five rows of the dataset with columns for the first action unit only.

Table 2.1: *data in wide format*

judge	video	frame	AU01-r	AU02-r	AU04-r	AU05-r	AU06-r	AU07-r	AU09-r	AU10-
Bell	McKell	1	0	0	0.69	0.63	0	1.5	0	0
Bell	McKell	2	0	0	0.69	0.63	0	1.5	0	0

The data can also be expressed in the long format with action unit being another index and presence and intensity being two columns. The Table 2.2 presents the first five rows of the data in the long format.

Table 2.2: *data in long format*

judge	video	frame	speaker	AU	presence	intensity
Bell	McKell	1	Appellent	AU01	1	0.00
Bell	McKell	1	Appellent	AU02	1	0.00
Bell	McKell	1	Appellent	AU04	0	0.69
Bell	McKell	1	Appellent	AU05	1	0.63
Bell	McKell	1	Appellent	AU06	0	0.00
Bell	McKell	1	Appellent	AU07	1	1.54
Bell	McKell	1	Appellent	AU09	1	0.00
Bell	McKell	1	Appellent	AU10	1	0.00
Bell	McKell	1	Appellent	AU12	0	0.00
Bell	McKell	1	Appellent	AU14	0	0.00
Bell	McKell	1	Appellent	AU15	1	0.00
Bell	McKell	1	Appellent	AU17	0	0.00
Bell	McKell	1	Appellent	AU20	1	0.05
Bell	McKell	1	Appellent	AU23	0	0.00
Bell	McKell	1	Appellent	AU25	1	0.00
Bell	McKell	1	Appellent	AU26	0	0.26
Bell	McKell	1	Appellent	AU28	NA	NA
Bell	McKell	1	Appellent	AU45	0	0.47
Bell	McKell	2	Appellent	AU01	0	0.00
Bell	McKell	2	Appellent	AU02	1	0.00
Bell	McKell	2	Appellent	AU04	1	0.69
Bell	McKell	2	Appellent	AU05	1	0.63
Bell	McKell	2	Appellent	AU06	0	0.00
Bell	McKell	2	Appellent	AU07	1	1.54
Bell	McKell	2	Appellent	AU09	1	0.00
Bell	McKell	2	Appellent	AU10	1	0.00
Bell	McKell	2	Appellent	AU12	0	0.00
Bell	McKell	2	Appellent	AU14	0	0.00
Bell	McKell	2	Appellent	AU15	1	0.00
Bell	McKell	2	Appellent	AU17	0	0.00
Bell	McKell	2	Appellent	AU20	1	0.05
Bell	McKell	2	Appellent	AU23	0	0.00
Bell	McKell	2	Appellent	AU25	1	0.00
Bell	McKell	2	Appellent	AU26	0	0.26
Bell	McKell	2	Appellent	AU28	NA	NA
Bell	McKell	2	Appellent	AU45	0	0.47

2.4 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.

2.5 Data cleaning

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

Method

3.1 Notation

Let $\mathbf{X}$ be a matrix of predictors, and $\mathbf{Y}$ 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

- X_1 indicates `judge` with six categories $i = 1, 2, \cdots, 6$
- X_2 indicates `video` for each of the seven cases, $j = 1, 2, \cdots, 7$
- X_3 indicates action unit containing 18 possible facial expression.
- X_4 indicates `speaker`, either the appellant or respondent, $l = 1, 2$
- X_5 indicates `frame` corresponding to time, $t = 1, 2, \cdots, T_j$

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.

A full, main effects model for the data might be expressed as:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl}$$

and we would be interested in interactions between judge, case, action unit and speaker. An alternative model structure, is to treat each action unit individually, and fit separate models.

Also, let P_{jitkl} represent the response variable presence, and I_{jitkl} represent the second response variable intensity. This notation will be helpful for defining the plots and models explained in this section.

3.2 Modelling

3.2.1 Model 1: action unit

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, action units and their interactions. The use of interaction terms allow for the effect of judge to be differed at different action unit level. The model can be written down as Equation 3.1. Judge Edelman and AU01 are selected as the base level.

$$P_{ik} = \mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik} \tag{3.1}$$

Based on this model structure, I can understand whether different factor level of action units have different effects on the same judges via ANOVA test. Analysis of Variance (ANOVA) test [reference] allows us to see if collectively, different factor levels of judge and action unit are the same or different ($\alpha_1 = \alpha_2 = \cdots = \alpha_6$). However, it doesn't explain which factor level causes this difference and thus we need multiple comparison.

3.2.2 Model 2 - Video

The second model as shown in Equation 3.2 is estimated to understand the interaction effect between judge and video while taking into account the main effect of judge, video and action unit and other pair wise interactions.

$$P_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \tag{3.2}$$

Building upon the previous model, we incorporate the effect of videos in this model. There are three main effects of judge, video and action unit in the model. We also incorporate the interaction term between judge and video, which allows the effect of judge to change at each video level. The interaction term for video and action unit is also added because this allows different videos to have different present score for each different action unit.

3.2.3 Model 3 - Appellant & respondent

The third model as shown in Equation 3.3 is estimated to understand the interaction effect between judge and speaking party.

$$P_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\delta)_{il} \tag{3.3}$$

3.3 Interaction effect

using only the main effect has long been criticised (**martinez2015analysing**)

In our context, the inclusion of more interactions will be helpful to explain the Presence score. However, the number of interactions is bounded by the degree of freedom we have. We will need to select the most important interactions to include while leave the less important ones.

3.4 Estimated Marginal Mean (EMM)

Estimated Marginal Mean (Sometimes called least square mean) is the prediction from a linear model over a defined reference grid.

- see the calculation of emm when presented with unbalanced data: weight is for averaging over a continuous variable, which is irrelevant for our study -> dont have to adjust weight
- typically the tests and confidence intervals are asymptotic. Thus the df column for tabular results will be Inf.[have a look at the confidence interval for glm: `https://cran.r-project.org/web/packages/emmeans/vignettes/models.html`]

3.5 Multiple Comparisons

- a very important paper on `multcomp` package: **hothorn2008simultaneous** other packages do the same thing: `linearHypothesis`

Testing significance based on the usage of p-value has been long criticised for its interpretation. Researchers can erroneously conclude significance becuase of p-value being less than 0.05 without discussing the false positive/negative rate. [add reference: (Yates, 1951; Savage, 1957; Rozeboom, 1960; Gardner and Altman, 1986; Simon, 1986; Bulpitt, 1987)]. The correct interpretation for a p-value being less than 0.05 is that the estimated quantity are significant subject to 5% of false positive rate. This indicates, if 100 tests are conducted simultaneously, we would expect around 5 variables being significant purly due to randomness.

This is especially a problem in our context when comparing the intensity score for all the available (judge, video, AU, speaker) pairs. The judges are expected to be impartial while we believe they may have some facial expressions during the hearing because they could be unbeknown to the speakers. It would be an concern if we don't control the false positive rate, which occurs when the judges didn't expression a facial expression but we claim they do.

Bonferroni adjustment (Rupert Jr, 2012; Efron and Hastie, 2016) is a method to reduce the false positive rate at the cost of relatively cheap false negative rate. In the context of large scale (N) hypothesis tests, rather than concluding signficance at α% for each individual test, Bonferroni method adjusts the critical value to α/N% for each single test. In this sense, Bonferroni adjustment controls the family-wise error rate at level α.

In both methods, the critical value of 0.05 is adjusted based on the number of simultaneous test and an assumption of independent test is made. [need some reference on both adjustment: `http://www.biostathandbook.com/multiplecomparisons.html`]

A better method to present the same result is through reporting the confidence interval and then comparing if intervals overlap with each other [some reference on the preference of confidence interval for doing comparison].

Chapter 4

Results

4.1 Action unit: Presence

4.1.1 Mean presence

I first compute the average presence (P_{ik}) of each action unit for each judge as

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

This is then plotted in Figure 4.1 to give an overview of the presence of all the action units across all the judge. The order of action unit on the y axis is ranked by the average presence of all the judge. The five most frequent action units are highlighted in blue for each judge and summarised in Table 4.1

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

Table 4.1: *The five most commonly presented action unit for each judge.*

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

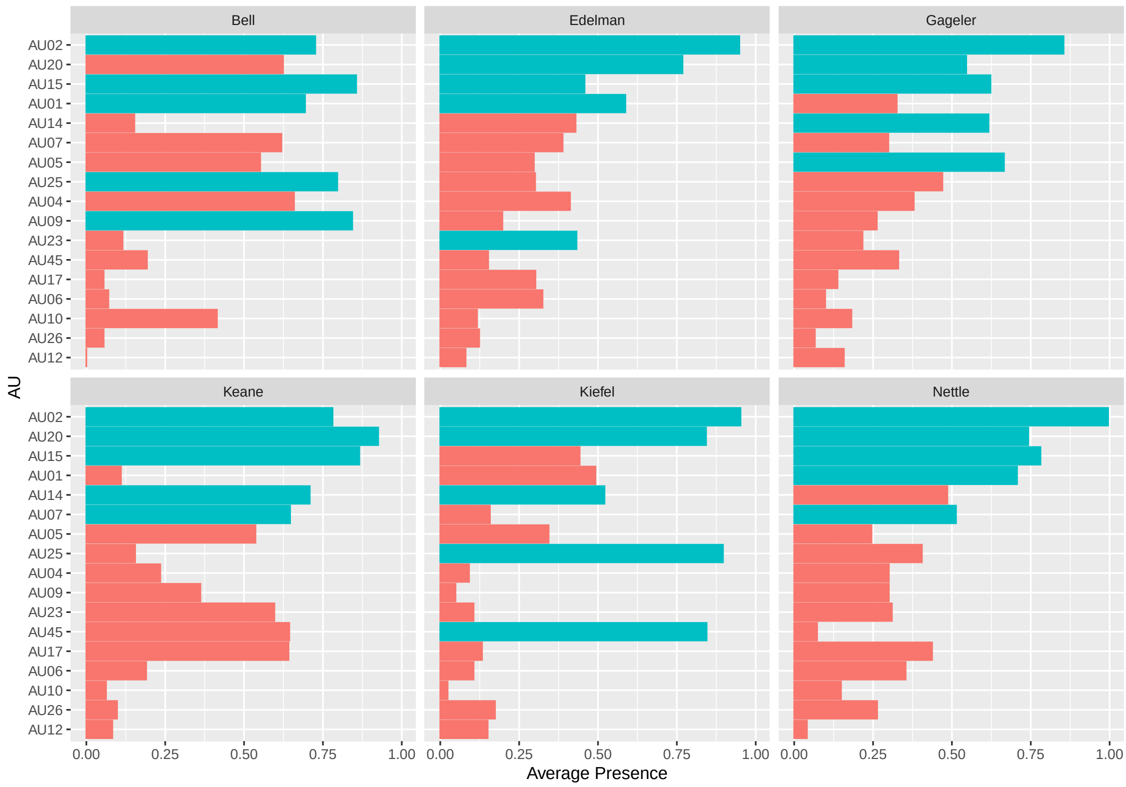


Figure 4.1: *The average presence score of each action unit for each judge, aggregating on video and time.*

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

According to Ekman, Friesen, and Hager (2002), AU02 makes a contribution to surprise, which is a positive attitude showing that judges are interested in a particular moment. AU14 indicates boredom and AU15 shows confusion. Along with other action units that presented with high frequency in a particular judge but not all (summarised in Table 4.3), the emotions judges displayed in the courtroom can be summarised into three categories, described in Table 4.2 along with the featured action units.

4.1.2 Interaction between judge and action unit

Analysis of Variance (ANOVA) test result in Table A.3 shows that there are significant variance for both judge, AU and their interactions and thus it is necessary to perform

Table 4.2: *Summarised emotions and featured action units*

emotion	Featured Action Unit
Surprise	AU01, AU02, AU05
Boredom	AU14, AU23
Confusion	AU07, AU15, AU23

Table 4.3: *Other high frequent action units*

judge	first	second	third
Bell	AU09: Nose wrinkler	AU25: Lips part	AU01: Inner brow raiser
Edelman	AU01: Inner brow raiser	AU23: Lip tightener	NA
Gageler	AU05: Upper lid raiser	NA	NA
Keane	AU07: Lid tightener	NA	NA
Kiefel	AU25: Lips part	AU45: Blink	NA
Nettle	AU01: Inner brow raiser	AU07: Lid tightener	NA

multiple comparison to determine how does each level of action unit differ for different judges. The estimated coefficient for each judge and AU pair is reported in Table 4.4 along with the upper and lower confidence interval bond. This result is also plotted in Figure 4.2 for visual inspection.

```
## [1] 0.0001276873
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  binomial_model_1$y, fitted(binomial_model_1)
## X-squared = 8.4887e-13, df = 8, p-value = 1
```

4.1.3 Presence by videos

Apart from visualising the general presence score for all the action units, I'm also interested in the break down statistics by video (P_{ijk}). This is computed as

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

for the four most common action units: AU02, AU14, AU15, AU20 and plotted in Figure 4.3. From this plot, we can observe that Gageler and Bell seems to have a larger fluctuation

Table 4.4: *model result*

judge	AU	prob	SE	df	asymp.LCL	asymp.UCL
Edelman	AU02	0.95	0.0069	Inf	0.92	0.97
Bell	AU02	0.73	0.0223	Inf	0.65	0.79
Gageler	AU02	0.85	0.0125	Inf	0.81	0.89
Keane	AU02	0.78	0.0207	Inf	0.71	0.84
Kiefel	AU02	0.95	0.0091	Inf	0.91	0.97
Nettle	AU02	1.00	0.0028	Inf	0.97	1.00
Edelman	AU14	0.43	0.0155	Inf	0.38	0.48
Bell	AU14	0.15	0.0179	Inf	0.10	0.22
Gageler	AU14	0.62	0.0173	Inf	0.56	0.67
Keane	AU14	0.71	0.0227	Inf	0.63	0.77
Kiefel	AU14	0.52	0.0211	Inf	0.46	0.58
Nettle	AU14	0.49	0.0201	Inf	0.42	0.55
Edelman	AU15	0.46	0.0156	Inf	0.41	0.51
Bell	AU15	0.86	0.0176	Inf	0.79	0.90
Gageler	AU15	0.62	0.0172	Inf	0.57	0.67
Keane	AU15	0.87	0.0170	Inf	0.80	0.91
Kiefel	AU15	0.44	0.0210	Inf	0.38	0.51
Nettle	AU15	0.78	0.0166	Inf	0.73	0.83
Edelman	AU20	0.77	0.0132	Inf	0.72	0.81
Bell	AU20	0.62	0.0242	Inf	0.55	0.69
Gageler	AU20	0.55	0.0177	Inf	0.49	0.60
Keane	AU20	0.93	0.0131	Inf	0.87	0.96
Kiefel	AU20	0.84	0.0154	Inf	0.79	0.88
Nettle	AU20	0.74	0.0176	Inf	0.68	0.79

across videos than others, especially for case OKS. This "visual" difference is only a preliminary result we observe from the plot, more solid statistical method need to be employed to validate the fluctuation shows statistical difference.

4.1.4 Interaction between judge and video

The estimated coefficients for the second model are presented in Table A.4 in the Appendix. Visual representation of the estimated result is shown in Figure 4.4. We can observe that Judge Edelman, Keane and Kiefel behave relatively consistent throughout all the videos since all the intervals for the same judge, same action unit but different videos overlaps after bonferroni adjustment. This would indicates in these videos, these judges would have similar thinking or reaction towards the evidence and argument presented by the barristers.

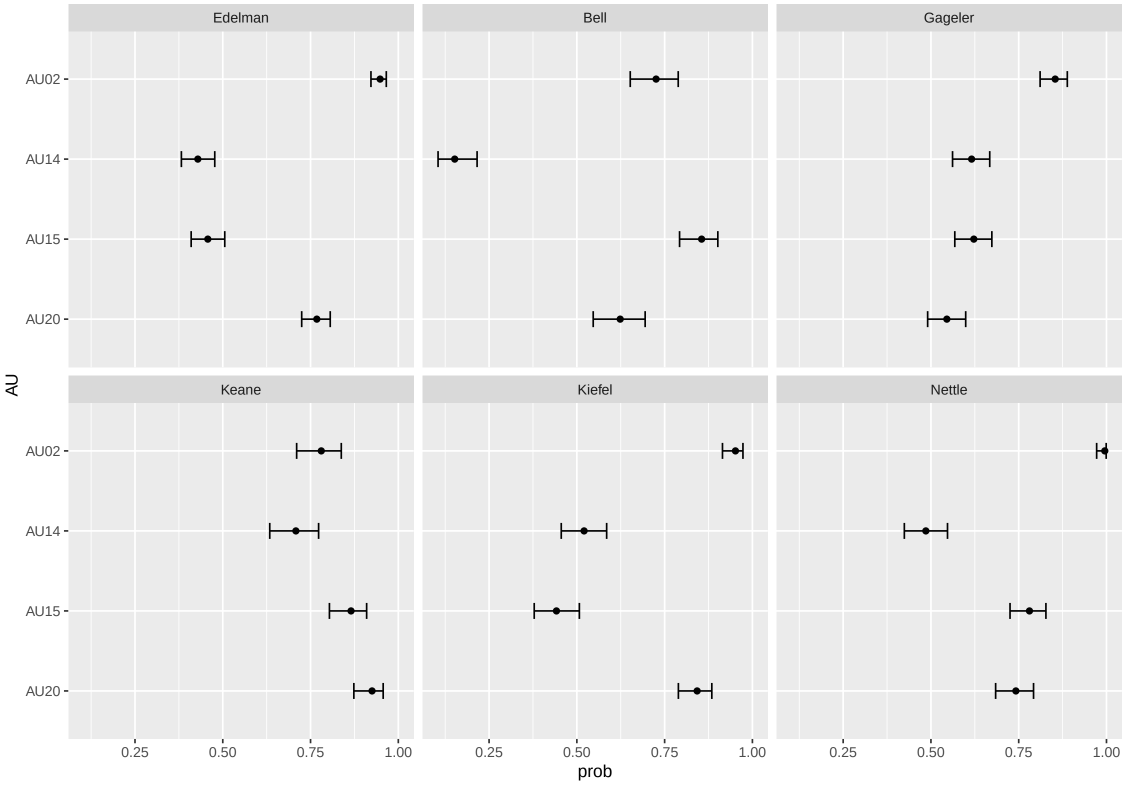


Figure 4.2: *THis is the graphical representation of model1*

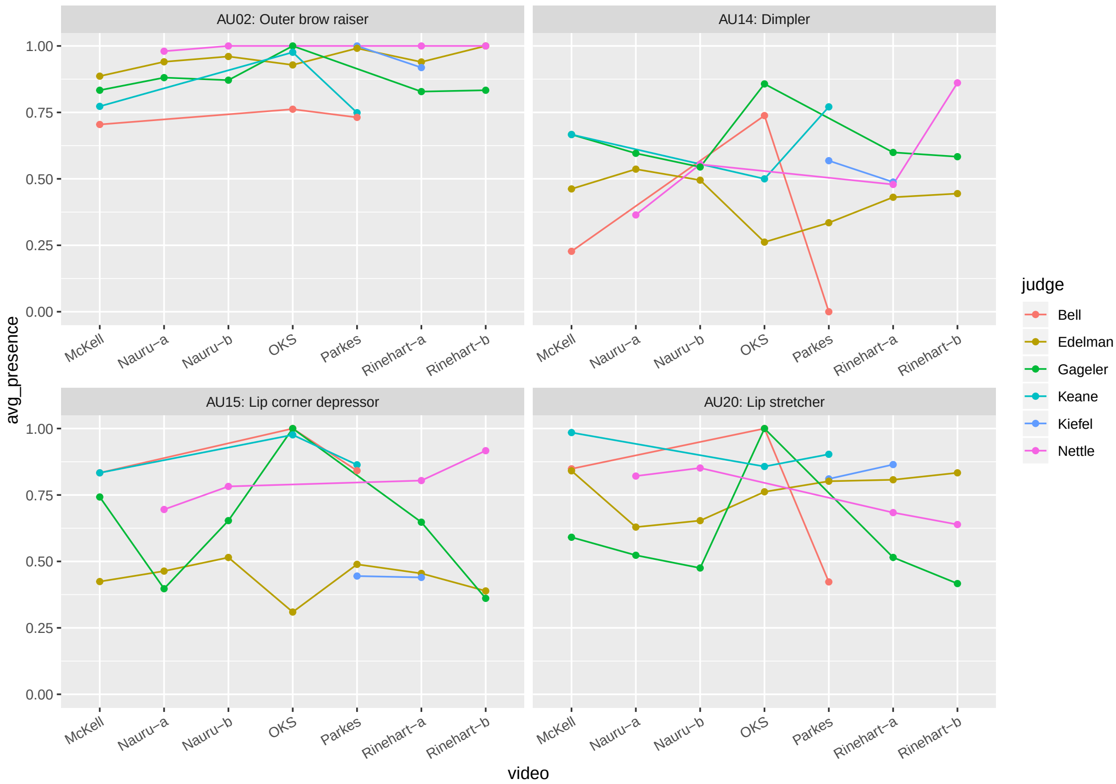


Figure 4.3: *Average presence of the four most common action units for each judge by video*

Judge Gageler seems to have a large fluctuate of his facial expressions in video OKS and his response is significantly different from those in other cases for AU15 and AU20. This shows consistency with our exploratory data analysis where Gageler tends to show a higher proportion of presence for action units in case OKS.

Our result validates the exploratory data analysis that the difference between case OKS and Parkes for Bell are significant in action unit 14, 15 and 20. Since the confidence interval doesn't overlap for these two videos. The difference between case McKell and OKS are not significance after the Bonferroni adjustment and this indicates the difference we see before in Figure 4.3 is more likely due to randomness rather than the true underlying difference between the two videos. From a legal perspective, this would show that Bell is addressing the cases with different responses. However, this different approach of responding by the judge doesn't indicate the biasness of the judge in the courtroom but the individuality of different judge approaching to cases.

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: presence
##
## Terms added sequentially (first to last)
##
##
##             Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                       15183      18900
## judge        5   298.54     15178      18602 < 2.2e-16 ***
## video        6    88.49     15172      18513 < 2.2e-16 ***
## AU           3  1693.00     15169      16820 < 2.2e-16 ***
## judge:video 13   181.90     15156      16638 < 2.2e-16 ***
## judge:AU    15  1148.69     15141      15490 < 2.2e-16 ***
## video:AU    18    97.25     15123      15392  7.04e-13 ***
```

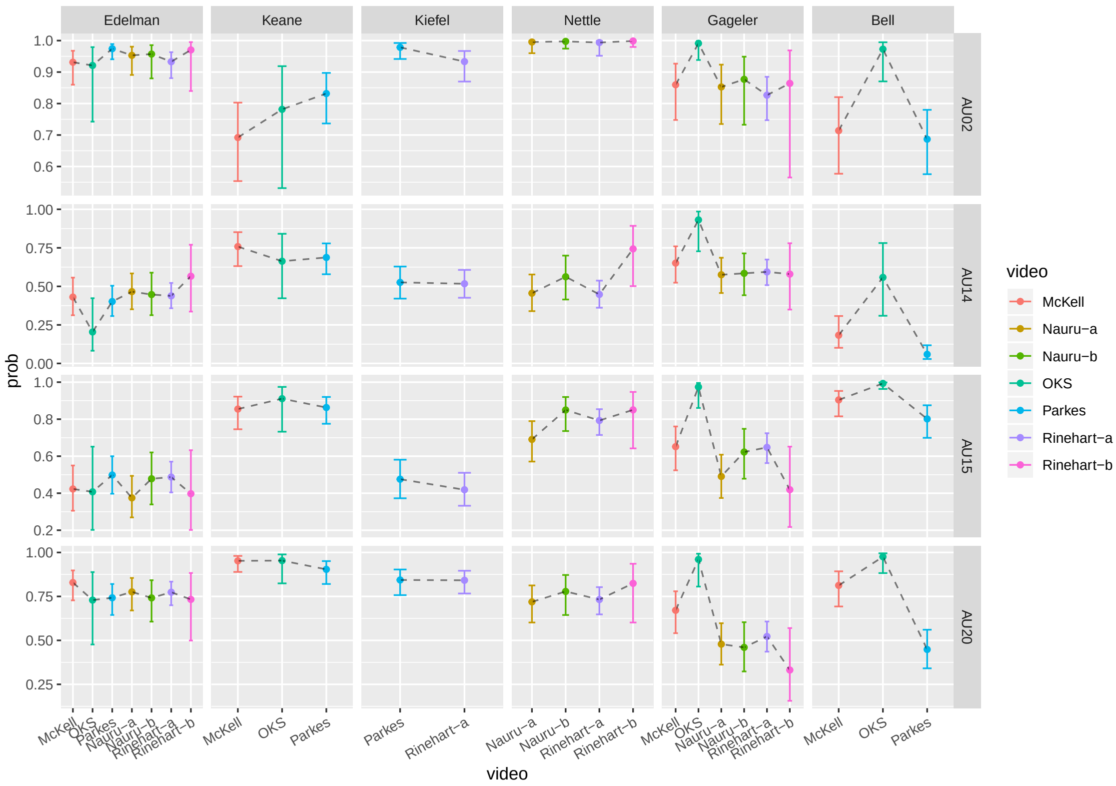


Figure 4.4: *This is the graphical representaion for model 2*

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] 0.06125016

##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  binomial_model_2$y, fitted(binomial_model_2)
## X-squared = 15.647, df = 8, p-value = 0.04773
```

4.1.5 Interaction between judge and speaking parties

The estimated coefficients are presented in Table A.5 and we could find from Figure 4.5 that the video-wise difference between judge still persist and the speaker-wise difference is not significant even before bomferroni adjustment. This result would be a validation that on the high court level, the judges are behaving impartial to different speaking parties.

Analysis of Deviance Table

Model: binomial, link: logit

Response: presence

Terms added sequentially (first to last)

```
          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
```

NULL 15183 18900

judge 5 298.54 15178 18602 < 2.2e-16 ***speaker 1 1.48 15177 18600 0.2237024***
video 6 87.14 15171 18513 < 2.2e-16 AU 3 1693.02 15168 16820 < 2.2e-16 ***judge:speaker 5 23.23 15163 16797 0.0003048*** judge:video 13 178.52 15150 16618 < 2.2e-16 ***judge:AU 15 1151.33 15135 15467 < 2.2e-16*** video:AU 18 97.35 15117 15370 6.75e-13 *** — Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ‘ ’ 1

```
## [1] 0.07359228
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  binomial_model_3$y, fitted(binomial_model_3)
## X-squared = 22.58, df = 8, p-value = 0.003948
```

4.1.6 Summary

- Based on the goodness-of-fit test, the p-value of binomial_model_3 is slightly greater than 0.05, indicating the model may be a good fit.
- However, the hosmer-lemeshow test shows binomial_model_1 has p-value = 1, which indicates model_1 is an adequate fit. other two models has p-value less than 0.05.

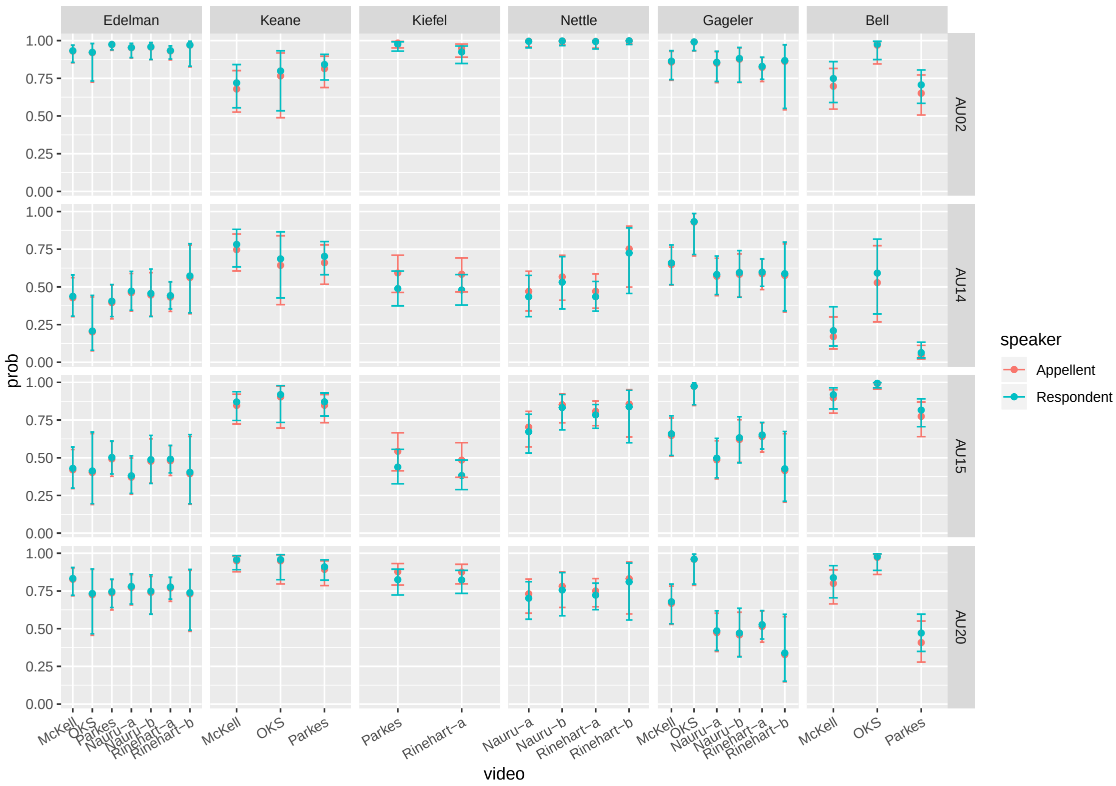


Figure 4.5: *THis is the graphical representation of model3*

- ANOVA tests suggest all the variables in model_1 and model_2 are significant individually, but speaker in model_3 is not significant at 5% significant level.

see if the three-way or four-way interaction works for multcomp package to produce simultaneous CI

4.2 Action unit: Intensity

4.2.1 General Intensity plot

In Ekman's 20002 FACS manual, the intensity of an action unit is defined based on five classes: Trace: 0-1, Slight: 1-2, Marked or pronounced: 2-3, Severe or extreme: 3-4 and Maximum: 4-5.

The boxplot of the intensity for all the judges across all the videos is presented in Figure 4.6. Each bar-and-whisker represents the intensity (I_{ijtk}) of all the action units aggregated on time for a particular judge i in a specific case j. For example, the first bar-and-whisker

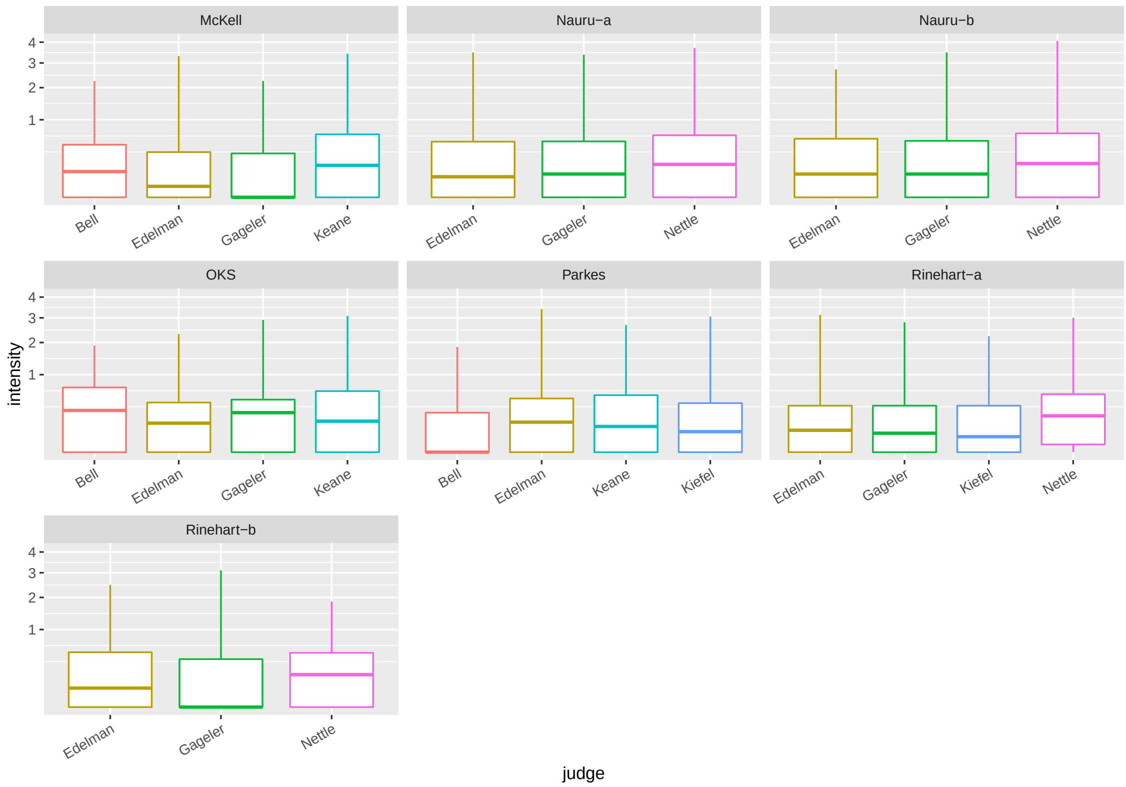


Figure 4.6: *General intensity score by judge and video*

in case Nauru_a is created using all the 17 action units of Edelman through out the elapsed time in Nauru_a case.

From the plot, we can see that most of the action units have low intensity score and this is expected because usually judges are expected to behave neutral in the court room. Thus a square root transformation is taken on the y axis for better visualisation effect. We can find that Judge Nettle seems to have higher average in all the four cases he appears: Nauru_a&b, Rinehart_a &b.

4.2.2 Mean intensity

Mean intensity score (I_{ik}) of each action unit for each of the judge is computed as

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

and plotted in Figure 4.7. The five most intense action units for each judge are presented in Table 4.5. We can find that the common high intense action units includes

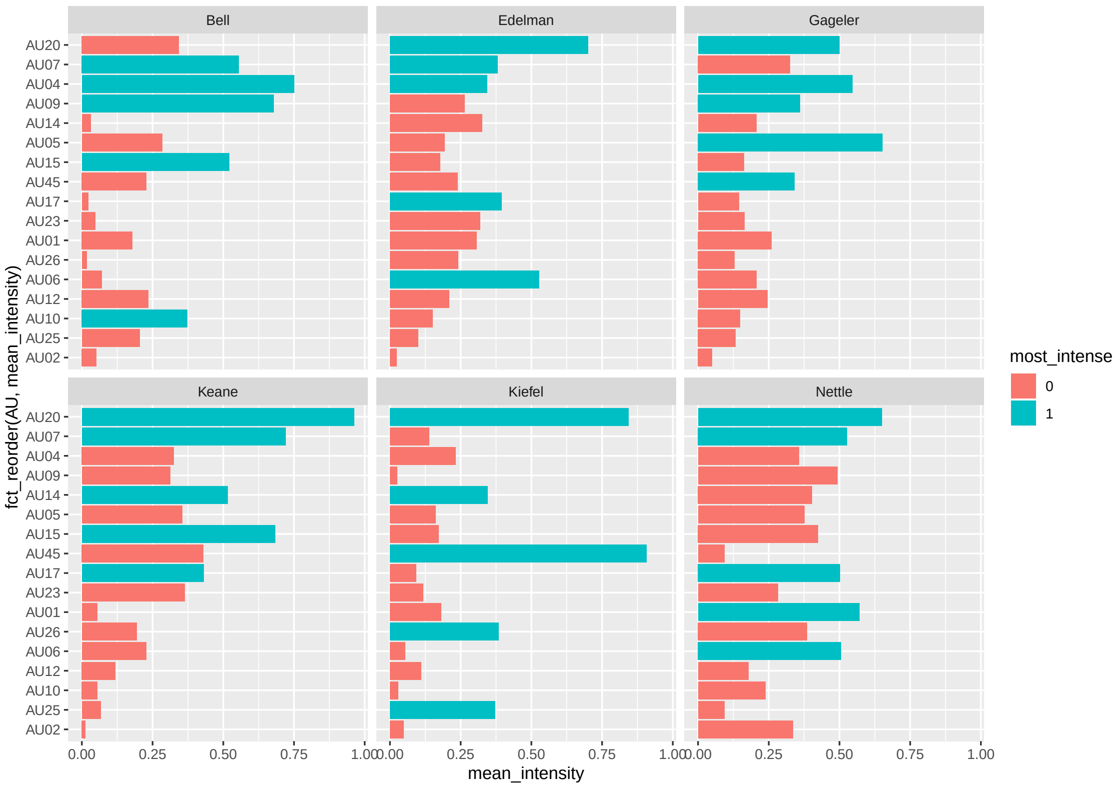


Figure 4.7: *Mean intensity score for each judge and action unit aggregating on videos.*

Table 4.5: *The five most intense action unit for each judge.*

index	Bell	Edelman	Gageler	Keane	Kiefel	Nettle
1	AU04	AU20	AU05	AU20	AU45	AU20
2	AU09	AU06	AU04	AU07	AU20	AU01
3	AU07	AU17	AU20	AU15	AU26	AU07
4	AU15	AU07	AU09	AU14	AU25	AU06
5	AU10	AU04	AU45	AU17	AU14	AU17

- AU20 (Lip Stretcher)
- AU07 (Lid Tightener)
- AU04 (Brow Lower)

AU04 also belongs to the confusion category as AU07. This could help to understand that judges are more likely to express a stronger confusing expression than other emotions.

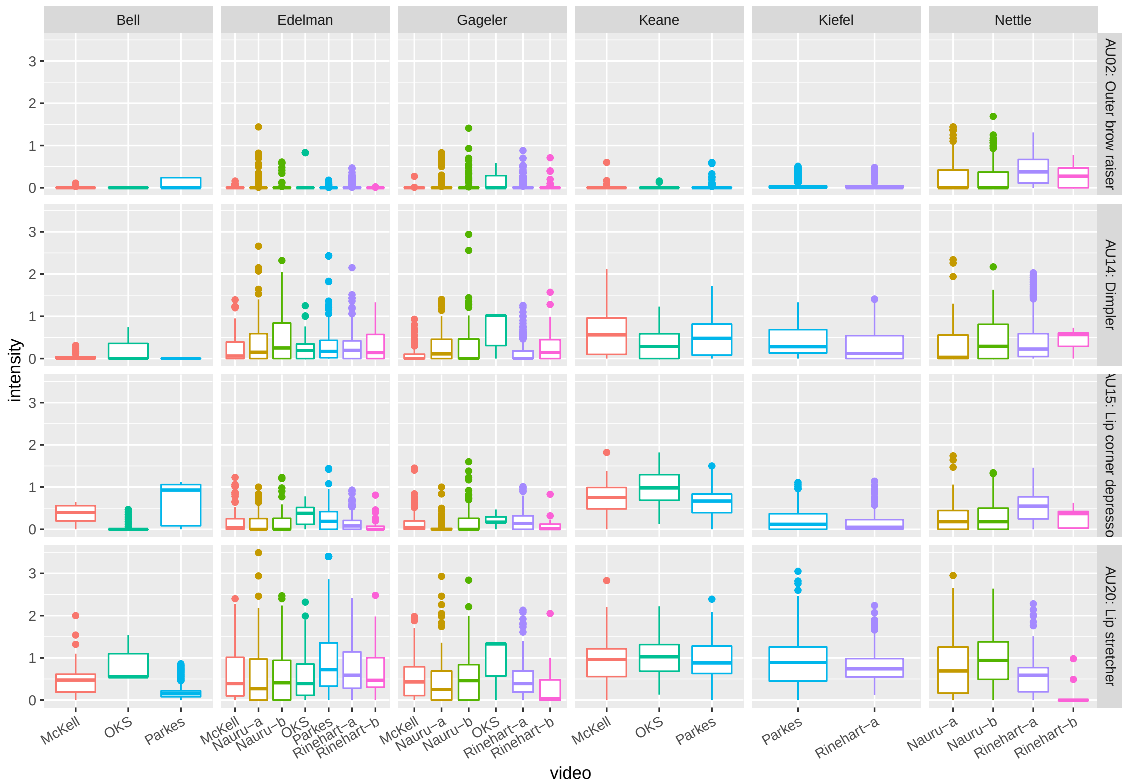


Figure 4.8: *Intensity score of the most frequent action units, seperating by judge and video ID.*

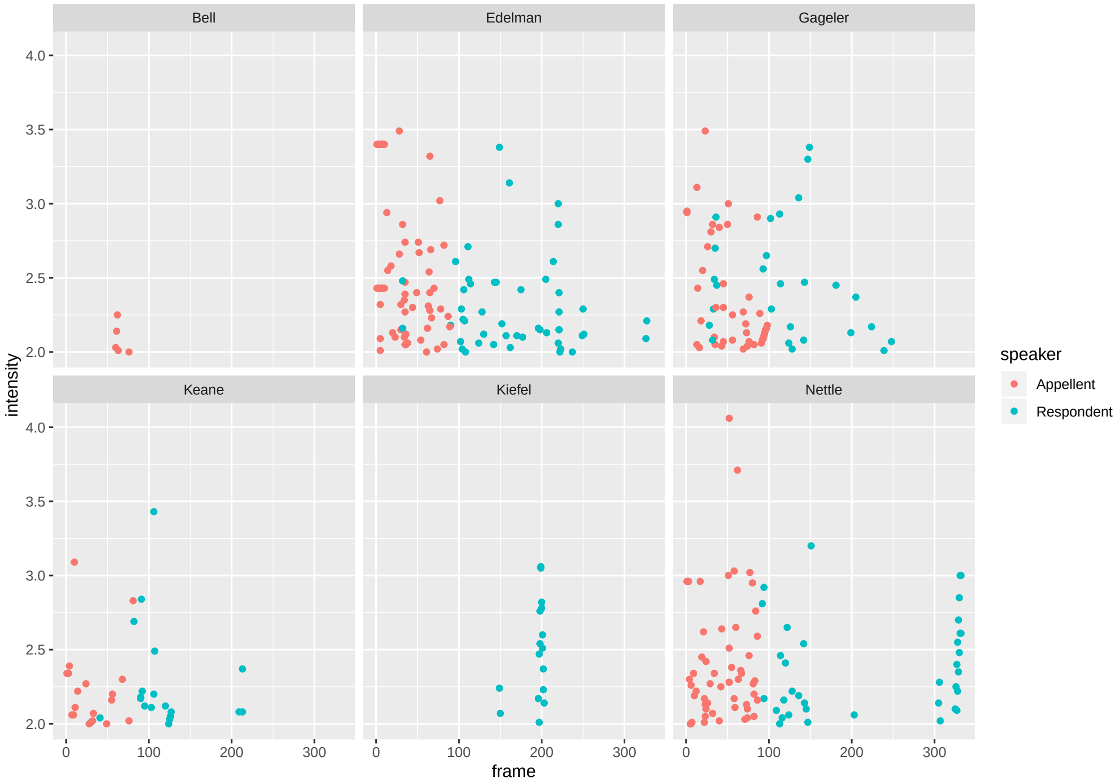
4.2.3 Model fit

4.2.4 Intensity plot for the most frequent action units

Apart from visualising the general intensity score for all the action units, I'm also interested in the intensity score of the most frequent action units. Figure 4.8 presents this. The statistics being plotted is I_{ijtk} with k including AU02, AU14, AU15 and AU20 as the most common four action units. From this plot, we can learn that AU02, although being commonly detected for all the judges, has low intensity score.

4.2.5 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.



Chapter 5

Conclusion

5.1 Discussion

5.2 Limitation

I will now briefly discuss some of the limitation of this work. The current image frames are extracted at every one minute interval. However, some facial expressions may only last for a few second. Thus more frequent time interval could be used for getting more precise facial information of the judges.

In my work, seven videos are being processed into the facial recognition software and more videos could be processed to get more robust results. The reason for not processing more videos in the current study is because the resolution of publicly available vidoes from the high court has only 720 pixels while the facial recognition software, OpenFace requires at least 30 pixels for a face to be detected. This means that we have to choose videos where three or five judges are presented.

However, this work has established a workflow for extracting facial expressions of human from videos. As long as more higher resolution videos are available, facial variables can be extracted via the same fashion.

5.3 Future work

Appendix A

Appendix

A.1 List of videos used in the project

Table A.1: *Information on the cases used in the project*

Case	Names	Link
The Republic of Nauru v WET040 [No. 2] [2018] HCA 60	Nauru-a	http://www
TTY167 v Republic of Nauru [2018] HCA 61	Nauru-b	http://www
Rinehart v Hancock Prospecting Pty Ltd [2019] HCA 13	Rinehart-a	http://www
Rinehart v Hancock Prospecting Pty Ltd [2019] HCA 13	Rinehart-b	http://www
Parkes Shire Council v South West Helicopters Pty Limited [2019] HCA 14	Parkes	ttp://www.
McKell v The Queen [2019] HCA 5	McKell	http://www
OKS v Western Australia [2019] HCA 10	OKS	http://www

A.2 List of the name of ction units

Table A.2: *The meaning of all the action unit estimated*

AU-meaning
AU01: Inner brow raiser
AU02: Outer brow raiser
AU04: Brow lowerer
AU05: Upper lid raiser
AU06: Cheek raiser
AU07: Lid tightener
AU09: Nose wrinkler
AU10: Upper lip raiser
AU12: Lip corner puller
AU14: Dimpler
AU15: Lip corner depressor
AU17: Chin raiser
AU20: Lip stretcher
AU23: Lip tightener
AU25: Lips part
AU26: Jaw drop
AU28: Lip suck
AU45: Blink

A.3 Model estimation result

Table A.4: *model result 2*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Edelman	McKell	AU02	0.931	0.01401	0.860	0.97
Bell	McKell	AU02	0.714	0.03411	0.577	0.82
Gageler	McKell	AU02	0.860	0.02419	0.748	0.93
Keane	McKell	AU02	0.692	0.03503	0.554	0.80
Edelman	Nauru-a	AU02	0.953	0.01127	0.891	0.98
Gageler	Nauru-a	AU02	0.853	0.02556	0.735	0.92

Table A.3: *ANOVA result*

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL	NA	NA	15183	18900.38	NA
judge	5	298.5383	15178	18601.84	0
AU	3	1682.9621	15175	16918.88	0
judge:AU	15	1113.8722	15160	15805.01	0

Table A.4: *model result 2*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Nettle	Nauru-a	AU02	0.995	0.00275	0.960	1.00
Edelman	Nauru-b	AU02	0.957	0.01263	0.880	0.99
Gageler	Nauru-b	AU02	0.877	0.02850	0.733	0.95
Nettle	Nauru-b	AU02	0.998	0.00161	0.974	1.00
Edelman	OKS	AU02	0.921	0.02810	0.743	0.98
Bell	OKS	AU02	0.973	0.01224	0.870	0.99
Gageler	OKS	AU02	0.991	0.00482	0.938	1.00
Keane	OKS	AU02	0.782	0.05426	0.531	0.92
Edelman	Parkes	AU02	0.974	0.00603	0.941	0.99
Bell	Parkes	AU02	0.687	0.02858	0.576	0.78
Keane	Parkes	AU02	0.832	0.02202	0.737	0.90
Kiefel	Parkes	AU02	0.979	0.00604	0.942	0.99
Edelman	Rinehart-a	AU02	0.933	0.01095	0.881	0.96
Gageler	Rinehart-a	AU02	0.827	0.01894	0.747	0.88
Kiefel	Rinehart-a	AU02	0.933	0.01271	0.870	0.97
Nettle	Rinehart-a	AU02	0.994	0.00353	0.952	1.00
Edelman	Rinehart-b	AU02	0.970	0.01455	0.840	1.00
Gageler	Rinehart-b	AU02	0.864	0.05155	0.565	0.97
Nettle	Rinehart-b	AU02	0.999	0.00095	0.980	1.00
Edelman	McKell	AU14	0.431	0.03440	0.313	0.56
Bell	McKell	AU14	0.183	0.02830	0.101	0.31
Gageler	McKell	AU14	0.651	0.03312	0.524	0.76
Keane	McKell	AU14	0.758	0.03063	0.631	0.85
Edelman	Nauru-a	AU14	0.466	0.03275	0.351	0.58
Gageler	Nauru-a	AU14	0.575	0.03218	0.457	0.69
Nettle	Nauru-a	AU14	0.456	0.03350	0.339	0.58
Edelman	Nauru-b	AU14	0.447	0.03911	0.313	0.59

Table A.4: *model result 2*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Gageler	Nauru-b	AU14	0.585	0.03848	0.442	0.71
Nettle	Nauru-b	AU14	0.562	0.04047	0.415	0.70
Edelman	OKS	AU14	0.204	0.04718	0.082	0.42
Bell	OKS	AU14	0.558	0.07074	0.309	0.78
Gageler	OKS	AU14	0.931	0.02873	0.728	0.99
Keane	OKS	AU14	0.663	0.06109	0.423	0.84
Edelman	Parkes	AU14	0.402	0.02753	0.307	0.50
Bell	Parkes	AU14	0.059	0.01160	0.029	0.12
Keane	Parkes	AU14	0.687	0.02796	0.579	0.78
Kiefel	Parkes	AU14	0.526	0.02919	0.421	0.63
Edelman	Rinehart-a	AU14	0.438	0.02287	0.358	0.52
Gageler	Rinehart-a	AU14	0.594	0.02316	0.508	0.67
Kiefel	Rinehart-a	AU14	0.517	0.02524	0.426	0.61
Nettle	Rinehart-a	AU14	0.448	0.02456	0.361	0.54
Edelman	Rinehart-b	AU14	0.566	0.06411	0.337	0.77
Gageler	Rinehart-b	AU14	0.580	0.06352	0.350	0.78
Nettle	Rinehart-b	AU14	0.743	0.05574	0.502	0.89
Edelman	McKell	AU15	0.423	0.03445	0.305	0.55
Bell	McKell	AU15	0.904	0.01816	0.815	0.95
Gageler	McKell	AU15	0.652	0.03329	0.524	0.76
Keane	McKell	AU15	0.855	0.02392	0.746	0.92
Edelman	Nauru-a	AU15	0.375	0.03152	0.269	0.49
Gageler	Nauru-a	AU15	0.491	0.03294	0.374	0.61
Nettle	Nauru-a	AU15	0.691	0.03057	0.571	0.79
Edelman	Nauru-b	AU15	0.478	0.03990	0.339	0.62
Gageler	Nauru-b	AU15	0.623	0.03813	0.478	0.75
Nettle	Nauru-b	AU15	0.850	0.02498	0.736	0.92

Table A.4: *model result 2*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Edelman	OKS	AU15	0.407	0.06677	0.202	0.65
Bell	OKS	AU15	0.993	0.00319	0.963	1.00
Gageler	OKS	AU15	0.974	0.01259	0.861	1.00
Keane	OKS	AU15	0.911	0.02957	0.732	0.97
Edelman	Parkes	AU15	0.498	0.02836	0.397	0.60
Bell	Parkes	AU15	0.801	0.02427	0.699	0.88
Keane	Parkes	AU15	0.863	0.01969	0.775	0.92
Kiefel	Parkes	AU15	0.476	0.02923	0.373	0.58
Edelman	Rinehart-a	AU15	0.487	0.02319	0.404	0.57
Gageler	Rinehart-a	AU15	0.648	0.02246	0.563	0.72
Kiefel	Rinehart-a	AU15	0.419	0.02488	0.332	0.51
Nettle	Rinehart-a	AU15	0.793	0.01925	0.715	0.85
Edelman	Rinehart-b	AU15	0.397	0.06348	0.202	0.63
Gageler	Rinehart-b	AU15	0.419	0.06417	0.218	0.65
Nettle	Rinehart-b	AU15	0.850	0.04053	0.642	0.95
Edelman	McKell	AU20	0.829	0.02334	0.728	0.90
Bell	McKell	AU20	0.813	0.02751	0.693	0.89
Gageler	McKell	AU20	0.671	0.03347	0.541	0.78
Keane	McKell	AU20	0.953	0.01143	0.889	0.98
Edelman	Nauru-a	AU20	0.776	0.02562	0.670	0.86
Gageler	Nauru-a	AU20	0.479	0.03319	0.362	0.60
Nettle	Nauru-a	AU20	0.719	0.02943	0.602	0.81
Edelman	Nauru-b	AU20	0.742	0.03287	0.607	0.84
Gageler	Nauru-b	AU20	0.460	0.03976	0.323	0.60
Nettle	Nauru-b	AU20	0.778	0.03157	0.644	0.87
Edelman	OKS	AU20	0.729	0.05918	0.477	0.89
Bell	OKS	AU20	0.976	0.01098	0.883	1.00

Table A.4: *model result 2*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Gageler	OKS	AU20	0.960	0.01853	0.806	0.99
Keane	OKS	AU20	0.954	0.01806	0.824	0.99
Edelman	Parkes	AU20	0.742	0.02448	0.645	0.82
Bell	Parkes	AU20	0.448	0.03085	0.341	0.56
Keane	Parkes	AU20	0.904	0.01730	0.821	0.95
Kiefel	Parkes	AU20	0.844	0.01992	0.758	0.90
Edelman	Rinehart-a	AU20	0.774	0.01857	0.700	0.83
Gageler	Rinehart-a	AU20	0.522	0.02394	0.436	0.61
Kiefel	Rinehart-a	AU20	0.842	0.01775	0.767	0.90
Nettle	Rinehart-a	AU20	0.732	0.02151	0.648	0.80
Edelman	Rinehart-b	AU20	0.733	0.05482	0.499	0.88
Gageler	Rinehart-b	AU20	0.331	0.06028	0.156	0.57
Nettle	Rinehart-b	AU20	0.824	0.04538	0.602	0.94

Table A.5: *model result 3*

judge	video	AU	speaker	prob	SE	asymp.LCL	asymp.UCL
Edelman	McKell	AU02	Appellent	0.930	0.0143	0.853	0.97
Bell	McKell	AU02	Appellent	0.697	0.0363	0.545	0.82
Gageler	McKell	AU02	Appellent	0.858	0.0247	0.737	0.93
Keane	McKell	AU02	Appellent	0.679	0.0371	0.526	0.80
Edelman	Nauru-a	AU02	Appellent	0.952	0.0115	0.884	0.98
Gageler	Nauru-a	AU02	Appellent	0.850	0.0263	0.722	0.93
Nettle	Nauru-a	AU02	Appellent	0.996	0.0026	0.958	1.00
Edelman	Nauru-b	AU02	Appellent	0.957	0.0127	0.874	0.99
Gageler	Nauru-b	AU02	Appellent	0.876	0.0286	0.722	0.95

Table A.5: *model result 3*

judge	video	AU	speaker	prob	SE	asymp.LCL	asymp.UCL
Nettle	Nauru-b	AU02	Appellent	0.998	0.0016	0.972	1.00
Edelman	OKS	AU02	Appellent	0.920	0.0287	0.724	0.98
Bell	OKS	AU02	Appellent	0.970	0.0137	0.845	0.99
Gageler	OKS	AU02	Appellent	0.991	0.0049	0.931	1.00
Keane	OKS	AU02	Appellent	0.766	0.0581	0.489	0.92
Edelman	Parkes	AU02	Appellent	0.974	0.0063	0.936	0.99
Bell	Parkes	AU02	Appellent	0.651	0.0358	0.506	0.77
Keane	Parkes	AU02	Appellent	0.814	0.0272	0.689	0.90
Kiefel	Parkes	AU02	Appellent	0.984	0.0048	0.951	0.99
Edelman	Rinehart-a	AU02	Appellent	0.931	0.0117	0.872	0.96
Gageler	Rinehart-a	AU02	Appellent	0.822	0.0209	0.729	0.89
Kiefel	Rinehart-a	AU02	Appellent	0.949	0.0106	0.890	0.98
Nettle	Rinehart-a	AU02	Appellent	0.994	0.0032	0.951	1.00
Edelman	Rinehart-b	AU02	Appellent	0.970	0.0148	0.825	1.00
Gageler	Rinehart-b	AU02	Appellent	0.862	0.0523	0.541	0.97
Nettle	Rinehart-b	AU02	Appellent	0.999	0.0009	0.978	1.00
Edelman	McKell	AU14	Appellent	0.427	0.0349	0.303	0.56
Bell	McKell	AU14	Appellent	0.170	0.0276	0.089	0.30
Gageler	McKell	AU14	Appellent	0.647	0.0339	0.511	0.76
Keane	McKell	AU14	Appellent	0.747	0.0326	0.605	0.85
Edelman	Nauru-a	AU14	Appellent	0.461	0.0337	0.339	0.59
Gageler	Nauru-a	AU14	Appellent	0.570	0.0334	0.442	0.69
Nettle	Nauru-a	AU14	Appellent	0.470	0.0355	0.341	0.60
Edelman	Nauru-b	AU14	Appellent	0.446	0.0392	0.306	0.59
Gageler	Nauru-b	AU14	Appellent	0.583	0.0386	0.434	0.72
Nettle	Nauru-b	AU14	Appellent	0.567	0.0405	0.411	0.71
Edelman	OKS	AU14	Appellent	0.201	0.0470	0.076	0.43

Table A.5: *model result 3*

judge	video	AU	speaker	prob	SE	asymp.LCL	asymp.UCL
Bell	OKS	AU14	Appellent	0.528	0.0733	0.268	0.77
Gageler	OKS	AU14	Appellent	0.930	0.0294	0.705	0.99
Keane	OKS	AU14	Appellent	0.643	0.0646	0.382	0.84
Edelman	Parkes	AU14	Appellent	0.395	0.0300	0.289	0.51
Bell	Parkes	AU14	Appellent	0.050	0.0109	0.022	0.11
Keane	Parkes	AU14	Appellent	0.660	0.0352	0.517	0.78
Kiefel	Parkes	AU14	Appellent	0.593	0.0332	0.463	0.71
Edelman	Rinehart-a	AU14	Appellent	0.432	0.0260	0.337	0.53
Gageler	Rinehart-a	AU14	Appellent	0.586	0.0268	0.482	0.68
Kiefel	Rinehart-a	AU14	Appellent	0.584	0.0301	0.467	0.69
Nettle	Rinehart-a	AU14	Appellent	0.471	0.0305	0.359	0.59
Edelman	Rinehart-b	AU14	Appellent	0.562	0.0646	0.322	0.78
Gageler	Rinehart-b	AU14	Appellent	0.576	0.0641	0.334	0.79
Nettle	Rinehart-b	AU14	Appellent	0.752	0.0549	0.498	0.90
Edelman	McKell	AU15	Appellent	0.419	0.0349	0.295	0.55
Bell	McKell	AU15	Appellent	0.897	0.0196	0.796	0.95
Gageler	McKell	AU15	Appellent	0.648	0.0340	0.511	0.76
Keane	McKell	AU15	Appellent	0.847	0.0255	0.724	0.92
Edelman	Nauru-a	AU15	Appellent	0.371	0.0323	0.258	0.50
Gageler	Nauru-a	AU15	Appellent	0.486	0.0341	0.360	0.61
Nettle	Nauru-a	AU15	Appellent	0.704	0.0315	0.572	0.81
Edelman	Nauru-b	AU15	Appellent	0.477	0.0400	0.332	0.63
Gageler	Nauru-b	AU15	Appellent	0.621	0.0383	0.470	0.75
Nettle	Nauru-b	AU15	Appellent	0.852	0.0247	0.732	0.92
Edelman	OKS	AU15	Appellent	0.403	0.0671	0.189	0.66
Bell	OKS	AU15	Appellent	0.992	0.0036	0.955	1.00
Gageler	OKS	AU15	Appellent	0.974	0.0129	0.846	1.00

Table A.5: *model result 3*

judge	video	AU	speaker	prob	SE	asymp.LCL	asymp.UCL
Keane	OKS	AU15	Appellent	0.903	0.0323	0.697	0.97
Edelman	Parkes	AU15	Appellent	0.492	0.0311	0.376	0.61
Bell	Parkes	AU15	Appellent	0.775	0.0303	0.640	0.87
Keane	Parkes	AU15	Appellent	0.848	0.0241	0.733	0.92
Kiefel	Parkes	AU15	Appellent	0.543	0.0340	0.414	0.67
Edelman	Rinehart-a	AU15	Appellent	0.480	0.0264	0.382	0.58
Gageler	Rinehart-a	AU15	Appellent	0.640	0.0259	0.538	0.73
Kiefel	Rinehart-a	AU15	Appellent	0.484	0.0309	0.370	0.60
Nettle	Rinehart-a	AU15	Appellent	0.808	0.0215	0.713	0.88
Edelman	Rinehart-b	AU15	Appellent	0.394	0.0636	0.191	0.64
Gageler	Rinehart-b	AU15	Appellent	0.415	0.0644	0.206	0.66
Nettle	Rinehart-b	AU15	Appellent	0.857	0.0395	0.639	0.95
Edelman	McKell	AU20	Appellent	0.827	0.0238	0.718	0.90
Bell	McKell	AU20	Appellent	0.800	0.0296	0.665	0.89
Gageler	McKell	AU20	Appellent	0.668	0.0342	0.528	0.78
Keane	McKell	AU20	Appellent	0.950	0.0122	0.877	0.98
Edelman	Nauru-a	AU20	Appellent	0.773	0.0265	0.658	0.86
Gageler	Nauru-a	AU20	Appellent	0.474	0.0343	0.348	0.60
Nettle	Nauru-a	AU20	Appellent	0.731	0.0301	0.603	0.83
Edelman	Nauru-b	AU20	Appellent	0.741	0.0330	0.598	0.85
Gageler	Nauru-b	AU20	Appellent	0.459	0.0398	0.316	0.61
Nettle	Nauru-b	AU20	Appellent	0.781	0.0314	0.641	0.88
Edelman	OKS	AU20	Appellent	0.725	0.0602	0.456	0.89
Bell	OKS	AU20	Appellent	0.973	0.0123	0.859	1.00
Gageler	OKS	AU20	Appellent	0.960	0.0190	0.787	0.99
Keane	OKS	AU20	Appellent	0.950	0.0198	0.797	0.99
Edelman	Parkes	AU20	Appellent	0.737	0.0267	0.624	0.83

Table A.5: *model result 3*

judge	video	AU	speaker	prob	SE	asymp.LCL	asymp.UCL
Bell	Parkes	AU20	Appellent	0.408	0.0368	0.279	0.55
Keane	Parkes	AU20	Appellent	0.893	0.0207	0.786	0.95
Kiefel	Parkes	AU20	Appellent	0.877	0.0183	0.790	0.93
Edelman	Rinehart-a	AU20	Appellent	0.769	0.0209	0.681	0.84
Gageler	Rinehart-a	AU20	Appellent	0.514	0.0276	0.411	0.62
Kiefel	Rinehart-a	AU20	Appellent	0.876	0.0168	0.797	0.93
Nettle	Rinehart-a	AU20	Appellent	0.750	0.0247	0.646	0.83
Edelman	Rinehart-b	AU20	Appellent	0.730	0.0554	0.482	0.89
Gageler	Rinehart-b	AU20	Appellent	0.328	0.0603	0.147	0.58
Nettle	Rinehart-b	AU20	Appellent	0.831	0.0443	0.598	0.94
Edelman	McKell	AU02	Respondent	0.933	0.0141	0.856	0.97
Bell	McKell	AU02	Respondent	0.749	0.0362	0.590	0.86
Gageler	McKell	AU02	Respondent	0.864	0.0246	0.741	0.93
Keane	McKell	AU02	Respondent	0.720	0.0386	0.554	0.84
Edelman	Nauru-a	AU02	Respondent	0.954	0.0112	0.887	0.98
Gageler	Nauru-a	AU02	Respondent	0.856	0.0258	0.729	0.93
Nettle	Nauru-a	AU02	Respondent	0.995	0.0030	0.952	1.00
Edelman	Nauru-b	AU02	Respondent	0.959	0.0125	0.875	0.99
Gageler	Nauru-b	AU02	Respondent	0.882	0.0287	0.724	0.95
Nettle	Nauru-b	AU02	Respondent	0.997	0.0018	0.967	1.00
Edelman	OKS	AU02	Respondent	0.923	0.0277	0.732	0.98
Bell	OKS	AU02	Respondent	0.976	0.0108	0.874	1.00
Gageler	OKS	AU02	Respondent	0.991	0.0047	0.934	1.00
Keane	OKS	AU02	Respondent	0.799	0.0526	0.534	0.93
Edelman	Parkes	AU02	Respondent	0.975	0.0060	0.939	0.99
Bell	Parkes	AU02	Respondent	0.707	0.0295	0.584	0.81
Keane	Parkes	AU02	Respondent	0.842	0.0222	0.739	0.91

Table A.5: *model result 3*

judge	video	AU	speaker	prob	SE	asymp.LCL	asymp.UCL
Kiefel	Parkes	AU02	Respondent	0.976	0.0069	0.930	0.99
Edelman	Rinehart-a	AU02	Respondent	0.934	0.0110	0.878	0.97
Gageler	Rinehart-a	AU02	Respondent	0.829	0.0192	0.744	0.89
Kiefel	Rinehart-a	AU02	Respondent	0.925	0.0144	0.848	0.96
Nettle	Rinehart-a	AU02	Respondent	0.994	0.0037	0.944	1.00
Edelman	Rinehart-b	AU02	Respondent	0.971	0.0142	0.830	1.00
Gageler	Rinehart-b	AU02	Respondent	0.868	0.0507	0.551	0.97
Nettle	Rinehart-b	AU02	Respondent	0.999	0.0010	0.974	1.00
Edelman	McKell	AU14	Respondent	0.438	0.0369	0.306	0.58
Bell	McKell	AU14	Respondent	0.210	0.0345	0.107	0.37
Gageler	McKell	AU14	Respondent	0.659	0.0353	0.515	0.78
Keane	McKell	AU14	Respondent	0.782	0.0330	0.633	0.88
Edelman	Nauru-a	AU14	Respondent	0.472	0.0348	0.345	0.60
Gageler	Nauru-a	AU14	Respondent	0.582	0.0343	0.450	0.70
Nettle	Nauru-a	AU14	Respondent	0.435	0.0370	0.303	0.58
Edelman	Nauru-b	AU14	Respondent	0.456	0.0429	0.303	0.62
Gageler	Nauru-b	AU14	Respondent	0.595	0.0423	0.431	0.74
Nettle	Nauru-b	AU14	Respondent	0.531	0.0477	0.354	0.70
Edelman	OKS	AU14	Respondent	0.208	0.0483	0.079	0.44
Bell	OKS	AU14	Respondent	0.591	0.0715	0.320	0.82
Gageler	OKS	AU14	Respondent	0.933	0.0282	0.715	0.99
Keane	OKS	AU14	Respondent	0.686	0.0612	0.427	0.87
Edelman	Parkes	AU14	Respondent	0.405	0.0284	0.303	0.52
Bell	Parkes	AU14	Respondent	0.064	0.0127	0.030	0.13
Keane	Parkes	AU14	Respondent	0.703	0.0293	0.581	0.80
Kiefel	Parkes	AU14	Respondent	0.489	0.0309	0.374	0.60
Edelman	Rinehart-a	AU14	Respondent	0.442	0.0240	0.354	0.53

Table A.5: *model result 3*

judge	video	AU	speaker	prob	SE	asymp.LCL	asymp.UCL
Gageler	Rinehart-a	AU14	Respondent	0.598	0.0242	0.504	0.69
Kiefel	Rinehart-a	AU14	Respondent	0.480	0.0271	0.379	0.58
Nettle	Rinehart-a	AU14	Respondent	0.435	0.0263	0.339	0.54
Edelman	Rinehart-b	AU14	Respondent	0.573	0.0652	0.328	0.79
Gageler	Rinehart-b	AU14	Respondent	0.588	0.0647	0.341	0.80
Nettle	Rinehart-b	AU14	Respondent	0.725	0.0601	0.456	0.89
Edelman	McKell	AU15	Respondent	0.430	0.0370	0.298	0.57
Bell	McKell	AU15	Respondent	0.919	0.0173	0.824	0.96
Gageler	McKell	AU15	Respondent	0.659	0.0354	0.515	0.78
Keane	McKell	AU15	Respondent	0.871	0.0244	0.747	0.94
Edelman	Nauru-a	AU15	Respondent	0.381	0.0336	0.264	0.51
Gageler	Nauru-a	AU15	Respondent	0.498	0.0353	0.367	0.63
Nettle	Nauru-a	AU15	Respondent	0.673	0.0345	0.531	0.79
Edelman	Nauru-b	AU15	Respondent	0.487	0.0436	0.329	0.65
Gageler	Nauru-b	AU15	Respondent	0.633	0.0416	0.466	0.77
Nettle	Nauru-b	AU15	Respondent	0.833	0.0304	0.686	0.92
Edelman	OKS	AU15	Respondent	0.413	0.0678	0.196	0.67
Bell	OKS	AU15	Respondent	0.994	0.0028	0.964	1.00
Gageler	OKS	AU15	Respondent	0.975	0.0123	0.852	1.00
Keane	OKS	AU15	Respondent	0.919	0.0277	0.734	0.98
Edelman	Parkes	AU15	Respondent	0.502	0.0292	0.393	0.61
Bell	Parkes	AU15	Respondent	0.816	0.0242	0.707	0.89
Keane	Parkes	AU15	Respondent	0.871	0.0197	0.777	0.93
Kiefel	Parkes	AU15	Respondent	0.438	0.0306	0.328	0.56
Edelman	Rinehart-a	AU15	Respondent	0.491	0.0243	0.400	0.58
Gageler	Rinehart-a	AU15	Respondent	0.652	0.0234	0.559	0.73
Kiefel	Rinehart-a	AU15	Respondent	0.382	0.0261	0.289	0.48

Table A.5: *model result 3*

judge	video	AU	speaker	prob	SE	asymp.LCL	asymp.UCL
Nettle	Rinehart-a	AU15	Respondent	0.785	0.0208	0.695	0.85
Edelman	Rinehart-b	AU15	Respondent	0.404	0.0651	0.196	0.65
Gageler	Rinehart-b	AU15	Respondent	0.427	0.0660	0.211	0.67
Nettle	Rinehart-b	AU15	Respondent	0.838	0.0443	0.600	0.95
Edelman	McKell	AU20	Respondent	0.833	0.0241	0.721	0.91
Bell	McKell	AU20	Respondent	0.838	0.0277	0.705	0.92
Gageler	McKell	AU20	Respondent	0.678	0.0354	0.533	0.80
Keane	McKell	AU20	Respondent	0.958	0.0109	0.891	0.98
Edelman	Nauru-a	AU20	Respondent	0.780	0.0265	0.664	0.86
Gageler	Nauru-a	AU20	Respondent	0.486	0.0355	0.355	0.62
Nettle	Nauru-a	AU20	Respondent	0.702	0.0333	0.563	0.81
Edelman	Nauru-b	AU20	Respondent	0.749	0.0349	0.596	0.86
Gageler	Nauru-b	AU20	Respondent	0.471	0.0441	0.313	0.64
Nettle	Nauru-b	AU20	Respondent	0.756	0.0382	0.585	0.87
Edelman	OKS	AU20	Respondent	0.734	0.0591	0.466	0.90
Bell	OKS	AU20	Respondent	0.979	0.0097	0.886	1.00
Gageler	OKS	AU20	Respondent	0.961	0.0182	0.795	0.99
Keane	OKS	AU20	Respondent	0.958	0.0167	0.825	0.99
Edelman	Parkes	AU20	Respondent	0.745	0.0249	0.640	0.83
Bell	Parkes	AU20	Respondent	0.471	0.0333	0.349	0.60
Keane	Parkes	AU20	Respondent	0.910	0.0169	0.822	0.96
Kiefel	Parkes	AU20	Respondent	0.825	0.0223	0.724	0.89
Edelman	Rinehart-a	AU20	Respondent	0.777	0.0191	0.696	0.84
Gageler	Rinehart-a	AU20	Respondent	0.527	0.0251	0.432	0.62
Kiefel	Rinehart-a	AU20	Respondent	0.823	0.0200	0.734	0.89
Nettle	Rinehart-a	AU20	Respondent	0.722	0.0233	0.626	0.80
Edelman	Rinehart-b	AU20	Respondent	0.739	0.0551	0.489	0.89

Table A.5: *model result 3*

judge	video	AU	speaker	prob	SE	asymp.LCL	asymp.UCL
Gageler	Rinehart-b	AU20	Respondent	0.339	0.0622	0.151	0.60
Nettle	Rinehart-b	AU20	Respondent	0.810	0.0494	0.558	0.94

Bibliography

Aliotta, JM (1987-1988). Combining Judges' Attributes and Case Characteristics: An Alternative Approach to Explaining Supreme Court Decisionmaking. *Judicature* **71**, 277.

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

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.

Cohn, JF, TS Kruez, I Matthews, Y Yang, MH Nguyen, MT Padilla, F Zhou, and F De la Torre (2009). Detecting depression from facial actions and vocal prosody. In: *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp.1–7.

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

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

Efron, B and T Hastie (2016). *Computer age statistical inference*. Vol. 5. Cambridge University Press.

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.

Facial Action Coding System (n.d.). https://www.paulekman.com/facial-action-coding-system/.

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/.

Huang, CL and YM Huang (1997). Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification. *Journal of Visual Communication and Image Representation* **8**(3), 278–290.

Huber, B, D McDuff, C Brockett, M Galley, and B Dolan (2018). Emotional Dialogue Generation Using Image-Grounded Language Models. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: ACM, pp.277:1–277:12. http://doi.acm.org/10.1145/3173574.3173851.

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.

Kapoor, A, Y Qi, and RW Picard (2003). Fully automatic upper facial action recognition. In: *IEEE International SOI Conference. Proceedings (Cat. No.03CH37443)*, pp.195–202.

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.

Kobayashi, H and F Hara (1992). Recognition of Six basic facial expression and their strength by neural network. In: *Proceedings IEEE International Workshop on Robot and Human Communication*, pp.381–386.

Koppen, PJ van and JT Kate (1984). Individual Differences in Judicial Behavior: Personal Characteristics and Private Law Decision-Making. *Law and Society Review* **18**(2), 225–247.

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.

Kulik, CT and MB Perry Elissa L.and Pepper (2003). Here Comes the Judge: The Influence of Judge Personal Characteristics on Federal Sexual Harassment Case Outcomes. *Law and Human Behavior* **27**, 69–86.

Lien, JJJ, T Kanade, JF Cohn, and CC Li (2000). Detection, tracking, and classification of action units in facial expression. *Robotics and Autonomous Systems* **31**(3), 131–146.

Lucey, P, JF Cohn, T Kanade, J Saragih, Z Ambadar, and I Matthews (2010). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp.94–101.

Nagel, SS (1962). Testing Relations Between Judicial Characteristics and Judicial Decision-Making. *Western Political Quarterly* **15**(3), 425–437. eprint: https://doi.org/10.1177/106591296201500301.

Nasir, M, A Jati, PG Shivakumar, S Nallan Chakravarthula, and P Georgiou (2016a). Multimodal and Multiresolution Depression Detection from Speech and Facial Landmark

Features. In: *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. AVEC '16. Amsterdam, The Netherlands: ACM, pp.43–50. http://doi.acm.org/10.1145/2988257.2988261.

Nasir, M, A Jati, PG Shivakumar, S Nallan Chakravarthula, and P Georgiou (2016b). 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.

Pan, X and AFdC Hamilton (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology* **109**(3), 395–417. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bjop.12290.

Rupert Jr, G (2012). *Simultaneous statistical inference*. Springer Science and Business Media.

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.

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).

Steffensmeier, D and CL Britt (2001). Judges' Race and Judicial Decision Making: Do Black Judges Sentence Differently? *Social Science Quarterly* **82**(4), 749–764. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/0038-4941.00057.

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.

Tong, Y, W Liao, and Q Ji (2007). Facial Action Unit Recognition by Exploiting Their Dynamic and Semantic Relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(10), 1683–1699.

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.

Welch, S, M Combs, and J Gruhl (1988). Do Black Judges Make a Difference? *American Journal of Political Science* **32**(1), 126–136.

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.