

Exploration of Facial Expression in Videos and Transcript of Legal Proceedings

Huize Zhang

05 May, 2019

Contents

Statement of the Topic	2
Motivation	2
Literature Review	2
Legal study from a behaviour perspective	2
Facial Recognition	3
Aim	3
Research Plan	3
Result	4
Data Processing	4
Data Quality	5
Exploratory Data Analysis for Nauru-a data	5
Text Analysis	6
Appendix	7
Reference	16

Statement of the Topic

Justices' decisions in the courtroom has been discussed by the legal literature. Gender, political views, religious background of both the Justices and parties in the case have been considered. This paper will explore the facial behaviour of the Justices during the hearings. Audio Visual(AV) recordings and case transcript will be utilised to predict the decisions of each Justice.

Motivation

Tutton, Mack, & Roach Anleu (2018) have attempted to take advantage of the AV technology, which is made available online by the High Court of Australia (Australia, 2019). This paper utilises visual inspection of the videos to highlight when Justices depart from the expected norms of the judicial conducts. 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 & Reid (2018) on professional tennis players during the Grand Slam matches. This shows the potential to predict the outcome of the trials based on Justices' demeanour utilising contemporary tools and emotion tagging techniques.

Literature Review

The discussion of the literature is broken into two parts. The first is an overview of the current work in the legal study to understand the behaviour of the Justices and the second part reviews the existing facial recognition and emotion tagging technology.

Legal study from a behaviour perspective

People have attempted to predict the decisions of the Justices for centuries. Chief Justices of Australia & Zealand (2017) presents the following in a 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 the judicial demeanour (Tutton et al., 2018) and (Goffman, 1956). Paul Ekman suggests that from a behavioural perspective, some facial and vocal inflections are often unbeknown to the speakers themselves (Ekman, O'Sullivan, Friesen, & Scherer, 1991). Many scholars have exploited this, 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, Kumar, Motwani, & Yeres, 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 the Justices' vote and thus the court outcome. These factors include the use of pleasant and unpleasant language by Black, Treul, Johnson, & Goldman (2011), the frequency and content of Justices' questions by Shullman (2004) and Johnson, Black, Goldman, & Treul (2009). A statistical method is used by Epstein, Landes, & Posner (2010) to include a regression analysis to the number of questions asked by the Justices to infer the winning party in the court.

Other scholars including (Chen, Halberstam, & Alan, 2016; Chen, Halberstam, Yu, & others, 2017, and @schubert1992observing) have studied the emotion of the Justices from vocal characteristics and suggest that these vocal characteristics, especially the perceived masculinity is strongly correlated with the court outcomes. Dietrich, Enos, & Sen (2019) uses a multilevel logistic model with random effect to suggest 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 the 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 the 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 et al. (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.

A more accurate approach could be employed to understand the emotion of the Justices in the court. The advancement in computer vision algorithm and software development make it possible for researchers to apply the facial recognition technology to a wide range of studies. Statistical analysis for the Justices' facial images has yet to be used by researchers to understand their decision making process.

Facial Recognition

An Anatomical analysis of facial action was conducted in 1976 by Ekman & Friesen (1976), the Facial Action Code (FAC) Ekman & Friesen (1978) was developed and has been further revised by Ekman, Friesen, & Hager (2002). This decomposition of facial muscles is widely used in scientific research. It was applied in competitive sports, specifically tennis by Kovalchik & Reid (2018) and 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 real world images was the focus of events like Audio/Visual Emotion Challenge (Schuller et al., 2011; Schuller, Valster, Eyben, Cowie, & Pantic, 2012) and Emotion Recognition In The Wild Challenge and Workshop (Dhall, Goecke, Joshi, Wagner, & Gedeon, 2013; Kahou et al., 2013).

Facial recognition software has also been implemented by DeepFace (Taigman, Yang, Ranzato, & Wolf, 2014) from Facebook, and FaceNet (Schroff, Kalenichenko, & Philbin, 2015) from Google. OpenFace (Baltrušaitis et al., 2018) is the first of its kind 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, & Morency (2016)), the OpenFace toolkit has been used in different social research studies including depression classification (Yang et al. (2016) and Nasir, Jati, Shivakumar, Nallan Chakravarthula, & Georgiou (2016)).

Aim

The aim of this study is to explore the possibility of using facial recognition technology to understand Justices' decision. There are four specific objectives to achieve.

- Read in video streams and convert into numerical data format.
- Perform data quality check: investigate video and data quality.
- Exploratory data analysis of the obtained dataset.
- Predicting the trial outcomes using deep learning models.

Research Plan

The following table details the work has been done so far in the project.

Table 1: Reserach Plan for semester 1

Timeline	Tasks
Week 1-2	Explore facial recognition APIs and background reading
Week 3	Obtain videos (Youtube-dl), crop into video frames (ffmpeg) and extract faces from video frames (ImageMagick)
Week 4	Obtain facial variables via OpenFace for one video: Nauru-a
Week 5	Preliminary exploratory data analysis on Nauru-a dataset
Week 6	Obtain the full dataset and perform data quality check
Week 7	Text analysis on the case transcripts, augment the dataset with appellant and respondent information
Week 8-9	Proposal writing and presentation preparation

Future plan for the project is listed below.

Table 2: Future Reserach Plan

Timeline	Tasks
May	In depth exploratory data analysis on the full dataset
June - August	Summarise Justices' behaviour in the courtroom: - Are the Justices behave the same or differently across the trails? - Would the Justices behave differently when respondent or appellant speak?
September	Modelling the decision of the Justices based on video and transcript data
October	Thesis writing and presentation preparation

Result

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 1 in the Appendix. Youtube-dl (Hsuan, Amine, & 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). Tiapan (Kobakian & 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, Lemstra, Randers-Pehrson, & Roucres, 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 OpenFace are individual 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**. There are 711 variables in the csv files produced by OpenFace for each face supplied and these variables can be classified into the following categories:

- Confidence
- Gaze variables

- Eye landmarking (**eye_lmk**) variables
- Pose variables
- Face landmarking (**face_lmk**) variables
- Action Unit variables

A sketch of the dataset is provided in the following table.

Table 3: Data Format

Video_id	Frame_id	Judge_id	Confidence	Gaze	Eye_lmk	Pose	Face_lmk	Action Unit
Nauru_a	1	Judge1						
Nauru_a	2	Judge1						
Nauru_a	3	Judge1						
...						
Nauru_a	151	Judge1						
Nauru_a	1	Judge2						
Nauru_a	2	Judge2						
Nauru_a	3	Judge2						
...						
Nauru_a	151	Judge2						
Nauru_a	1	Judge3						
Nauru_a	2	Judge3						
Nauru_a	3	Judge3						
...						
Nauru_a	151	Judge3						
Nauru_b	1	Judge1						

Data Quality

Data Quality check is performed after obtaining the full dataset. Missing rate for each Justice in each video is summarised in the Table ?? in the Appendix. For some Justices in some videos, the missing rate could be due to the resolution of the source video and special care will be taken in the data exploration and modelling.

Exploratory Data Analysis for Nauru-a data

Exploratory Data Analysis has been done for Nauru-a video to understand different kinds of variables provided by OpenFace.

Confidence

The confidence variable shows how OpenFace is confident about the detection of the face. From Figure 2, a majority of the faces are detected with a confidence rate higher than 0.7, which shows the effectiveness of the OpenFace software. Several confidence levels from 0 to 1 are selected with the cropped face images displayed in Figure 3,4,5,6 in the Appendix for readers to have a better understanding. From these figures, the score of the confidence level is more related to the angle that the Justice's face present in the images. A lower confidence score is still useful information indicating that there's a higher probability that the Justices is looking at the paperwork at that moment.

Gaze and Pose Variable

Gaze and pose variables provides very similar information on the direction the eyes are looking and the direct the head is positioning towards in these courtroom images. In Figure 7, both Justice Edelman and Justices Gageler are looking straight for most of the time while Justices Nettle tends to look towards right hand side. This could due to the fact that Justice Nettle usually sit on the left hand side of the courtroom according to the relavant etiquette.

Face Landmarking

Face Landmarking locates the key points include eyebrows, eye, nose, lip and chin on the face and a standard facial landmarking guide is shown in Figure 8 in the Appendix. The computed facial landmarking by OpenFace is also plotted in Figure 9 and 10 in the Appendix with Figure 11 being the origin image frame from the video. Notice that OpenFace indeed captures the shape of Justice Nettle on the left most of the image.

Action Unit

Facial Action Unit(AU), as a way to describe human facial expression, has been discussed in the literature review section. OpenFace produces 17 Action Units and Figure 12 in the Appendix shows the average AU detected for every Justices in the *Nauru-a* video. The most frequentist detected AU are AU2, AU17 and AU20, which stands for outer eye brow raising, chin raising and lip stretcher.

Text Analysis

Interruptions has been studied a lot by interactional sociolinguistics (Friedrich, 1972; Kennedy & Camden, 1983; Konakahara, 2015). Smith-Lovin and Brody(1989) argues that if a person wants to move a discussion toward issues he or she prefers to discuss, interrupting another participant is often an effective strategy. Similarly, in the courtroom, interrupted by the judges has a negative implication on the odd of winning the case (Johnson et al., 2009). Transcripts for each case are used to compute the number of time appellant, respondent and each of the Justices speak and visualisation is created in Figure 14 and 13 in the appendix. A higher number of time the appellant speak means the appellant is interrupted more often by the Justices.

From Figure 13, the Chief Justice will interrupt the barristers more often than other Justices. In Figure 14, it seems that for the four cases that have been decided up until now, three out of four cases have the party being interrupted less win the cases, which coincides with the current literature mentioned above.

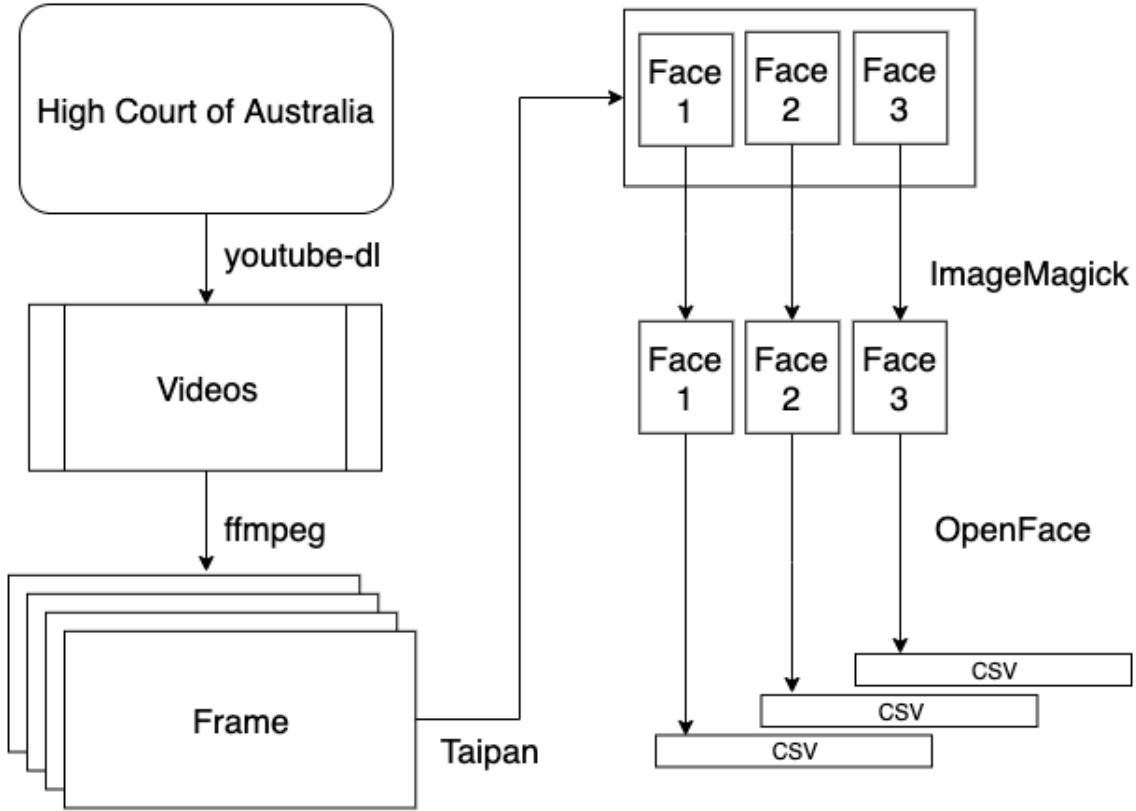


Figure 1: Workflow for data extraction

Appendix

Table 4: 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
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

Table 5: Missing rate for each Justice in all the videos

video_id	judge_id	na_count	count	na_prop	data_prop
Nauru_a	Edelman	3	151	0.02	0.98
Nauru_a	Gageler	3	151	0.02	0.98
Nauru_a	Nettle	3	151	0.02	0.98
Nauru_b	Edelman	0	101	0.00	1.00
Nauru_b	Gageler	0	101	0.00	1.00
Nauru_b	Nettle	0	101	0.00	1.00
McKell	Bell	114	132	0.86	0.14
McKell	Edelman	47	132	0.36	0.64
McKell	Gageler	63	132	0.48	0.52
McKell	Gordon	47	132	0.36	0.64
McKell	Keane	32	132	0.24	0.76
OKS	Bell	39	42	0.93	0.07
OKS	Edelman	26	42	0.62	0.38
OKS	Gageler	39	42	0.93	0.07
OKS	Gordon	42	42	1.00	0.00
OKS	Keane	13	42	0.31	0.69
Parkes	Bell	223	227	0.98	0.02
Parkes	Edelman	164	227	0.72	0.28
Parkes	Gordon	227	227	1.00	0.00
Parkes	Keane	63	227	0.28	0.72
Parkes	Kiefel	209	227	0.92	0.08
Rinehart_a	Edelman	213	332	0.64	0.36
Rinehart_a	Gageler	190	332	0.57	0.43
Rinehart_a	Gordon	330	332	0.99	0.01
Rinehart_a	Kiefel	305	332	0.92	0.08
Rinehart_a	Nettle	303	332	0.91	0.09
Rinehart_b	Edelman	21	36	0.58	0.42
Rinehart_b	Gageler	13	36	0.36	0.64
Rinehart_b	Gordon	36	36	1.00	0.00
Rinehart_b	Kiefel	36	36	1.00	0.00
Rinehart_b	Nettle	29	36	0.81	0.19

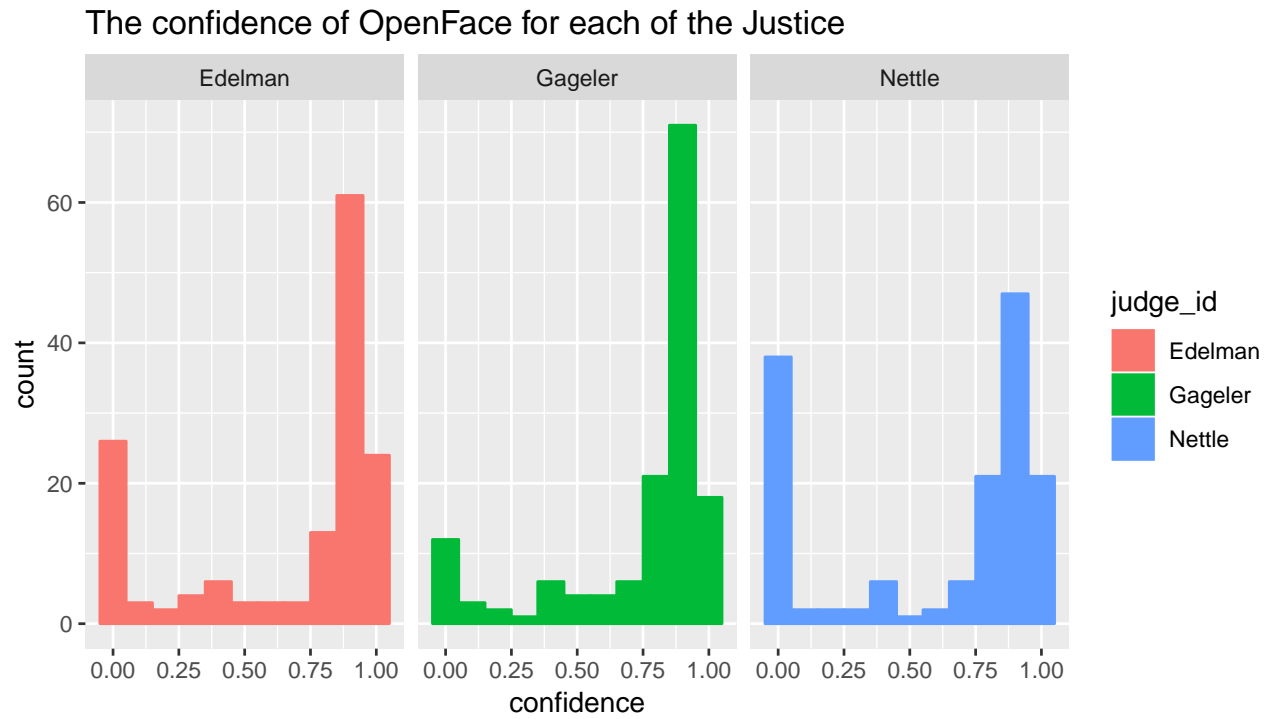


Figure 2: The confidence level of OpenFace for each of the three Justices in the Nauru-a video



Figure 3: Face with confidence = 0.925



Figure 4: Face with confidence = 0.775



Figure 5: Face with confidence = 0.425



Figure 6: Face with confidence = 0.025

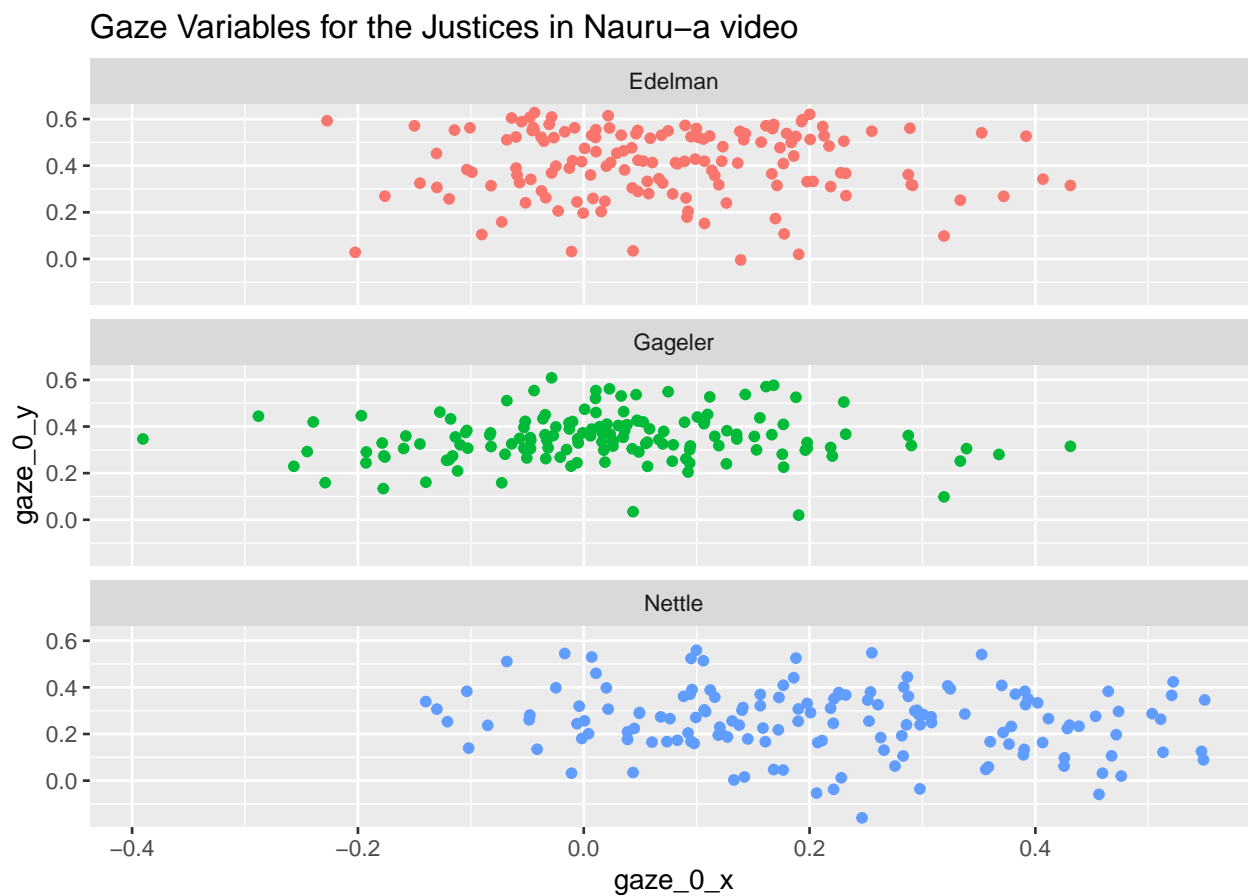


Figure 7: The direction the Justices' eyes are looking in each image frame.

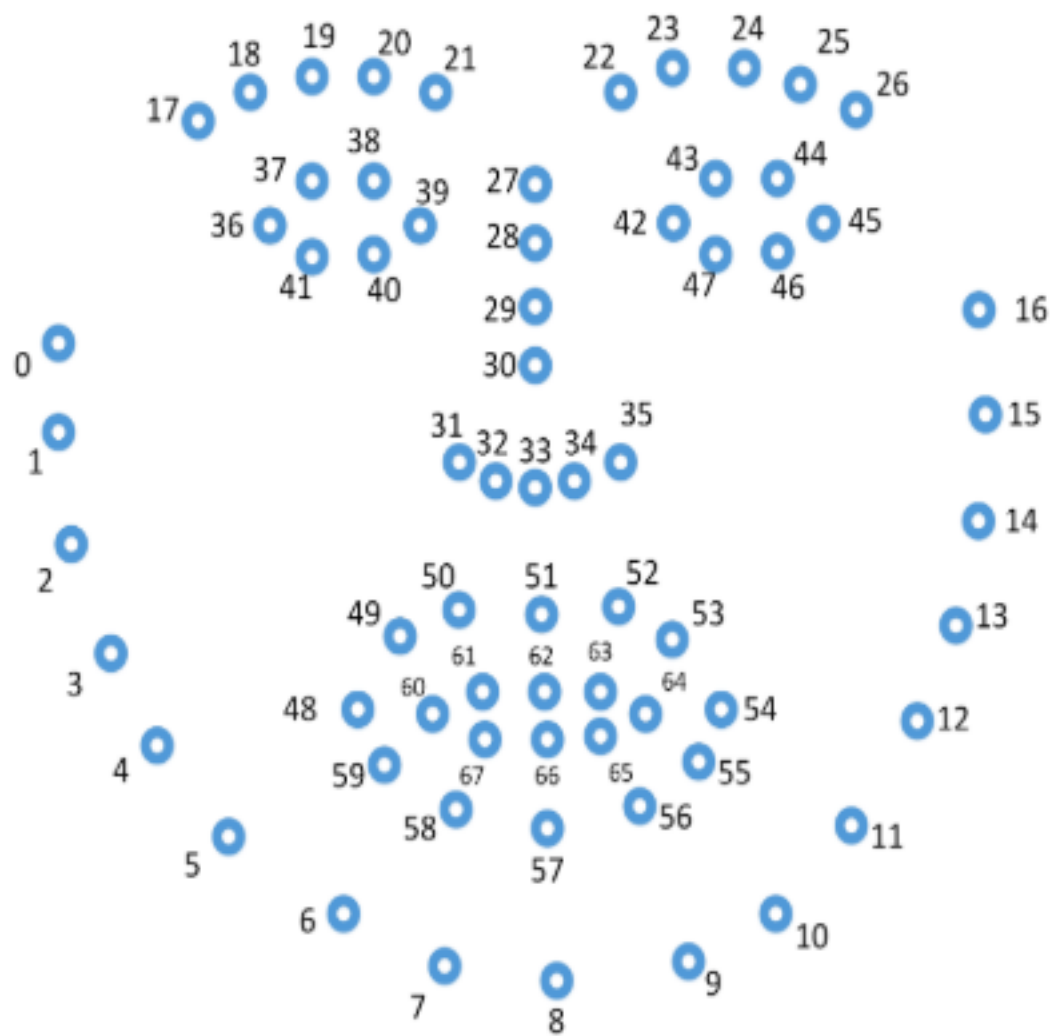


Figure 8: Position of key points in the face landmarking

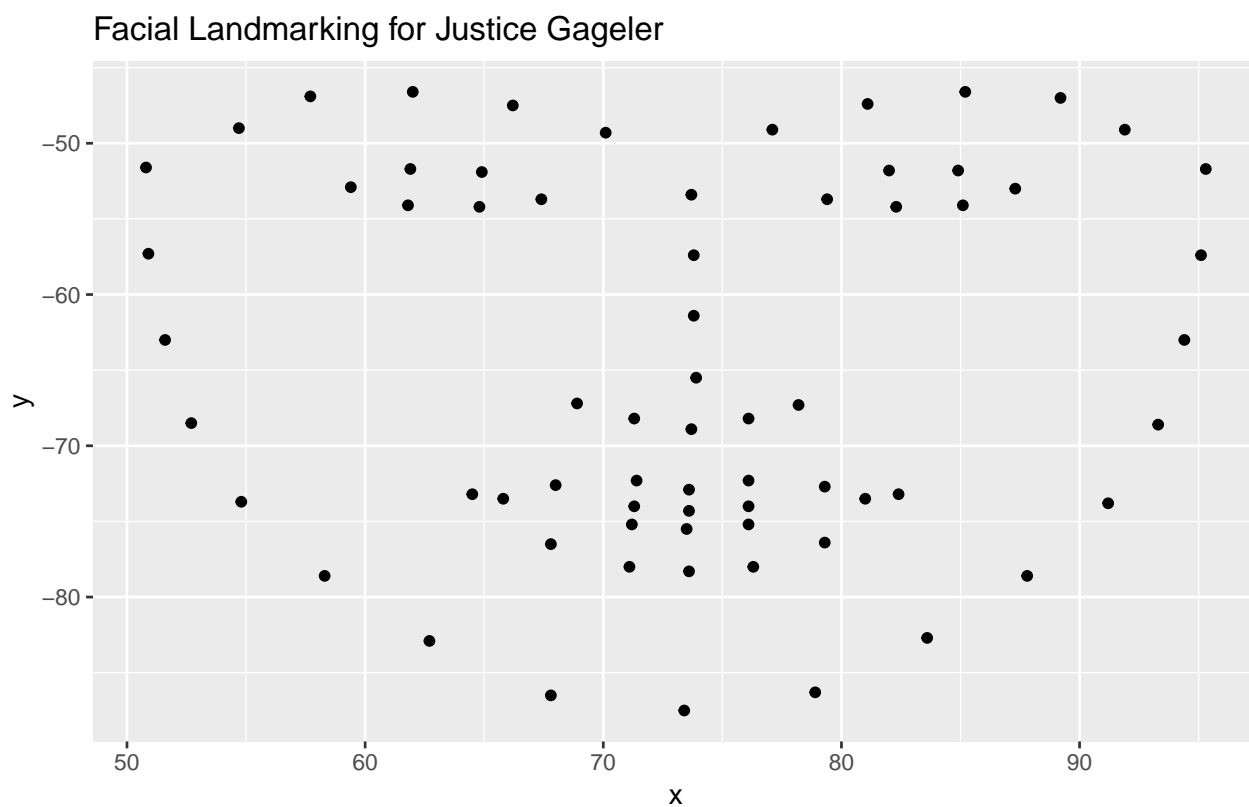


Figure 9: Facial Landmarking for Justice Gageler

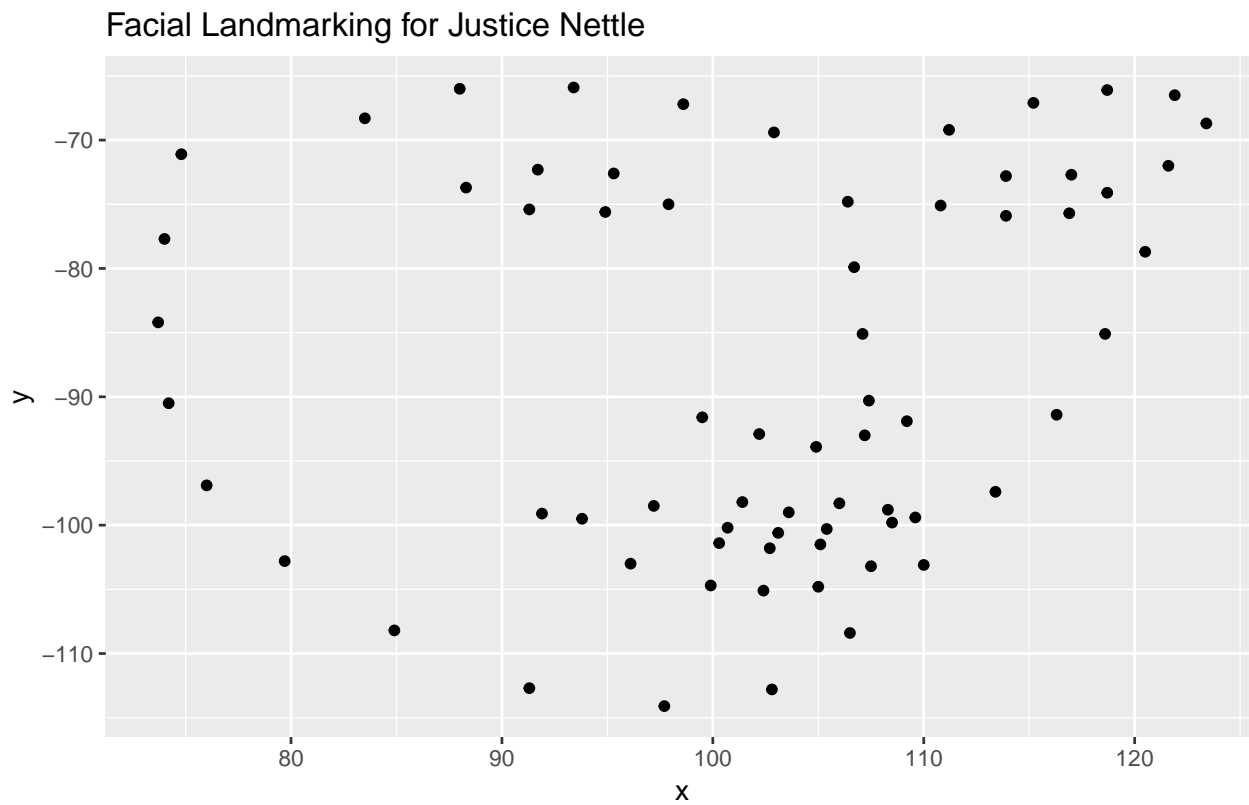


Figure 10: Facial Landmarking for Justice Nettle



Figure 11: Origin image from the Nauru-a video

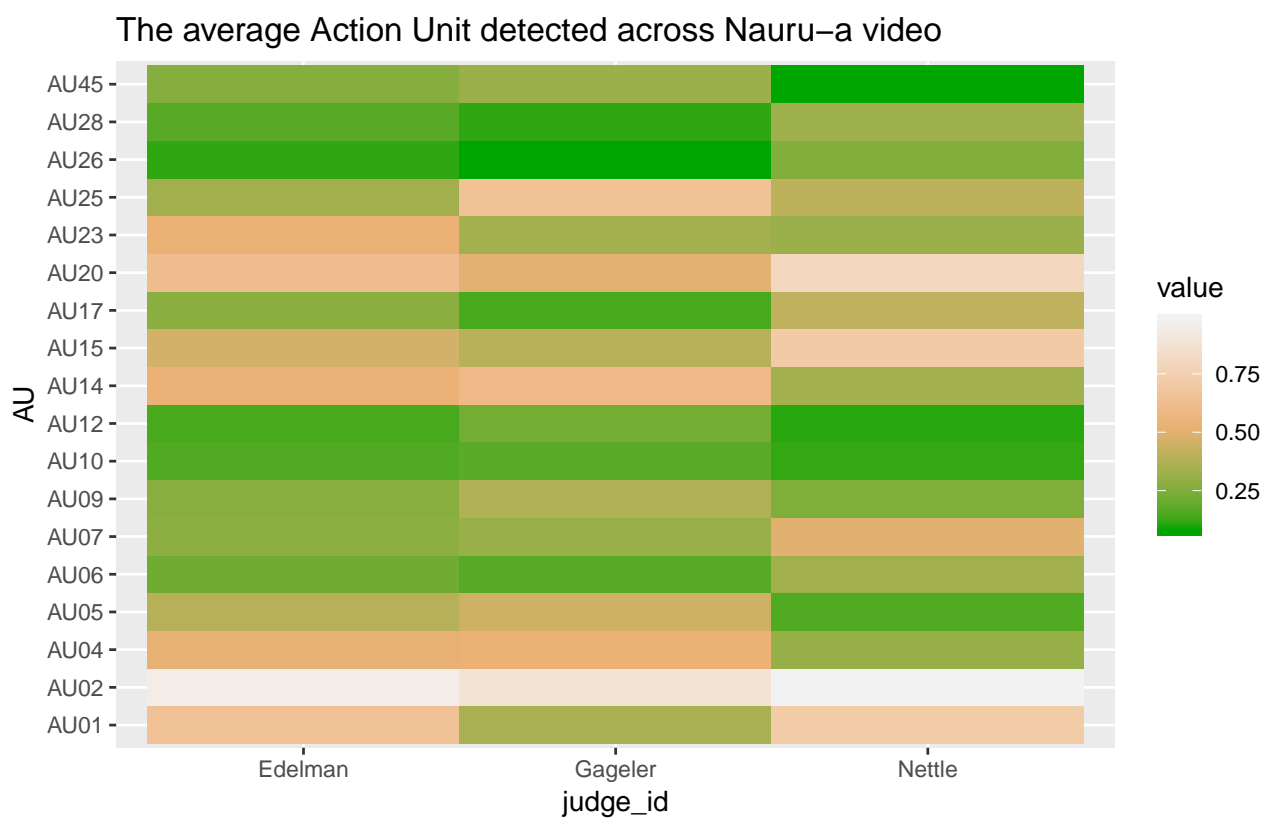


Figure 12: Facial Action Unit for three Justices in Nauru-a video

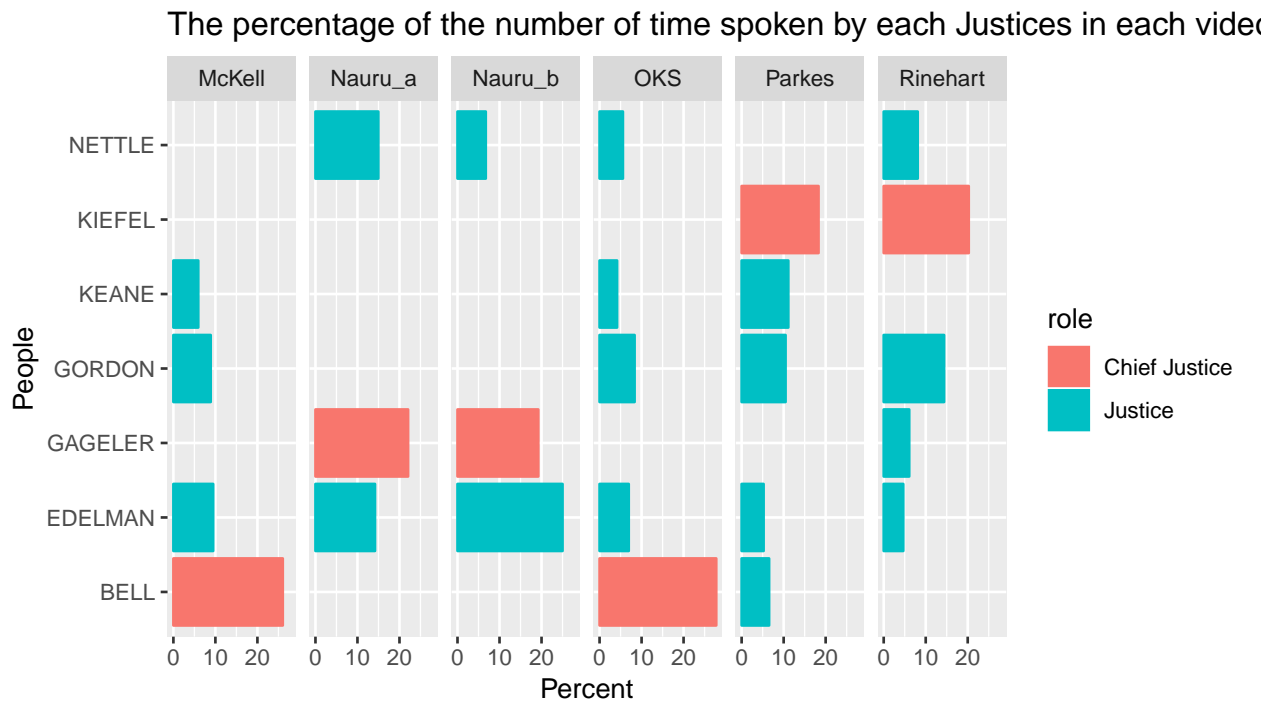


Figure 13: The number of time each Justices speak in each case

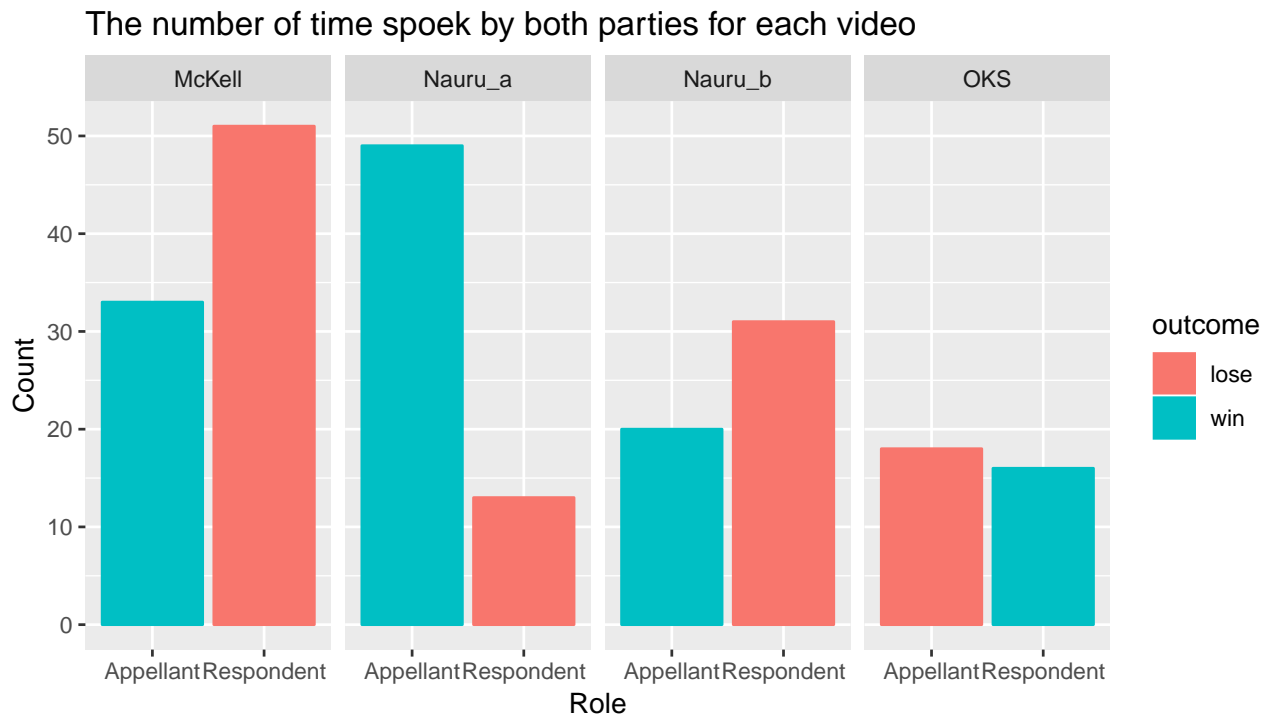


Figure 14: The number of time appellent and respondent speak for each case

Reference

- Australia, H. C. of. (2019). *Recent av recordings*. Retrieved from \url{http://www.hcourt.gov.au/cases/recent-av-recordings}
- Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. *2018 13th ieee international conference on automatic face & gesture recognition (fg 2018)*, 59–66. IEEE.
- Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2016). Openface: An open source facial behavior analysis toolkit. *2016 ieee winter conference on applications of computer vision (wacv)*, 1–10. IEEE.
- Bellard, F. (2019). *Ffmpeg*. Retrieved from <https://ffmpeg.org/>
- Black, R. C., Treul, S. A., Johnson, T. R., & Goldman, J. (2011). Emotions, oral arguments, and supreme court decision making. *The Journal of Politics*, 73(2), 572–581.
- Chen, D., Halberstam, Y., & Alan, C. (2016). Perceived masculinity predicts us supreme court outcomes. *PloS One*, 11(10), e0164324.
- Chen, D., Halberstam, Y., Yu, A., & others. (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).
- Chen, D. L., Kumar, M., Motwani, V., & Yeres, P. (2018). *Is justice really blind? And is it also deaf*. Technical report.
- Chief Justices of Australia, T. C. of, & Zealand, N. (2017). *Guide to judicial conduct* (3rd ed.). Melbourne: Austral-asian Institute of Judicial Administration.
- Cristy, J., Lemstra, D., Randers-Pehrson, G., & Roucres, B. (2019). *ImageMagick-dl*. Retrieved from <https://github.com/ImageMagick>
- Dhall, A., Goecke, R., Joshi, J., Wagner, M., & Gedeon, T. (2013). Emotion recognition in the wild challenge 2013. *Proceedings of the 15th acm on international conference on multimodal interaction*, 509–516. ACM.
- Dietrich, B. J., Enos, R. D., & Sen, M. (2019). Emotional arousal predicts voting on the us supreme court. *Political Analysis*, 27(2), 237–243.
- Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, 1(1), 56–75.
- Ekman, P., Friesen, W. V., & Hager, J. C. (2002). Facial action coding system: The manual on cd rom. *A Human Face, Salt Lake City*, 77–254.
- Ekman, P., O’Sullivan, M., Friesen, W. V., & Scherer, K. R. (1991). Invited article: Face, voice, and body in detecting deceit. *Journal of Nonverbal Behavior*, 15(2), 125–135.
- Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Palo Alto: CA: Consulting Psychologists Press.
- Epstein, L., Landes, W. M., & Posner, R. A. (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.
- Friedrich, P. (1972). Social context and semantic feature: The russian pronominal usage. *Directions in Sociolinguistics: The Ethnography of Communication*, 270, 287–295.
- Goffman, E. (1956). The nature of deference and demeanor. *American Anthropologist*, 58(3), 473–502.
- Hsuan, Y. C., Amine, R., & Sergey, M. (2019). *Youtube-dl*. Retrieved from <https://github.com/yt-dl-org/youtube-dl/>
- Johnson, T. R., Black, R. C., Goldman, J., & Treul, S. A. (2009). Inquiring minds want to know: Do justices tip their hands with questions at oral argument in the us supreme court. *Wash. UJL & Pol’y*, 29, 241.

- Kahou, S. E., Pal, C., Bouthillier, X., Froumenty, P., Gülçehre, Ç., Memisevic, R., ... others. (2013). Combining modality specific deep neural networks for emotion recognition in video. *Proceedings of the 15th acm on international conference on multimodal interaction*, 543–550. ACM.
- Kennedy, C. W., & Camden, C. (1983). Interruptions and nonverbal gender differences. *Journal of Nonverbal Behavior*, 8(2), 91–108.
- Kobakian, S., & O'Hara-Wild, M. (2018). *Taipan: Tool for annotating images in preparation for analysis*. Retrieved from <https://CRAN.R-project.org/package=taipan>
- Konakahara, M. (2015). An analysis overlapping questions in casual elf conversation: Cooperative or competitive contribution. *Journal of Pragmatics*, 84, 37–53.
- Kovalchik, S., & Reid, M. (2018). *Going inside the inner game: Predicting the emotions of professional tennis players from match broadcasts*. MIT Sloan Sports Analytics Conference. Accessible at [http://www ...](http://www...)
- Nasir, M., Jati, A., Shivakumar, P. G., Nallan Chakravarthula, S., & Georgiou, P. (2016). Multimodal and multiresolution depression detection from speech and facial landmark features. *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 43–50. ACM.
- Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. *Proceedings of the ieee conference on computer vision and pattern recognition*, 815–823.
- Schubert, J. N., Peterson, S. A., Schubert, G., & Wasby, S. (1992). Observing supreme court oral argument: A biosocial approach. *Politics and the Life Sciences*, 11(1), 35–52.
- Schuller, B., Valstar, M., Eyben, F., McKeown, G., Cowie, R., & Pantic, M. (2011). Avec 2011—the first international audio/visual emotion challenge. *International conference on affective computing and intelligent interaction*, 415–424. Springer.
- Schuller, B., Valster, M., Eyben, F., Cowie, R., & Pantic, M. (2012). AVEC 2012: The continuous audio/visual emotion challenge. *Proceedings of the 14th acm international conference on multimodal interaction*, 449–456. ACM.
- Shullman, S. L. (2004). The illusion of devil's advocacy: How the justices of the supreme court foreshadow their decisions during oral argument. *J. App. Prac. & Process* 2, 271(6).
- Smith-Lovin, L., & Brody, C. (1989). Interruptions in group discussions: The effects of gender and group composition. *American Sociological Review*, 424–435.
- Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. *Proceedings of the ieee conference on computer vision and pattern recognition*, 1701–1708.
- Tutton, J., Mack, K., & Roach Anleu, S. (2018). Judicial demeanor: Oral argument in the high court of australia. *Justice System Journal*, 39(3), 273–299.
- Yang, L., Jiang, D., He, L., Pei, E., Oveneke, M. C., & Sahli, H. (2016). Decision tree based depression classification from audio video and language information. *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 89–96. ACM.