What Goes On in the Courtroom? Audiovisual Analysis of Ninth Circuit Oral Arguments

Aniket Kesari, Elliott Ash, & Christoph Gösssman
Fordham University/ETH Zurich
Conference on Empirical Legal Studies
10/14/2023

Research Question + Data

- What Goes on in the Courtroom?
  - Do judges convey information through body language and speech?
  - Can this information be extracted at scale by a computer?
- Data
  - ~10,000 videos downloaded from Ninth Circuit's YouTube Channel
  - LexisNexis data with judge names, appointing President, votes, case characteristics

Previous Literature

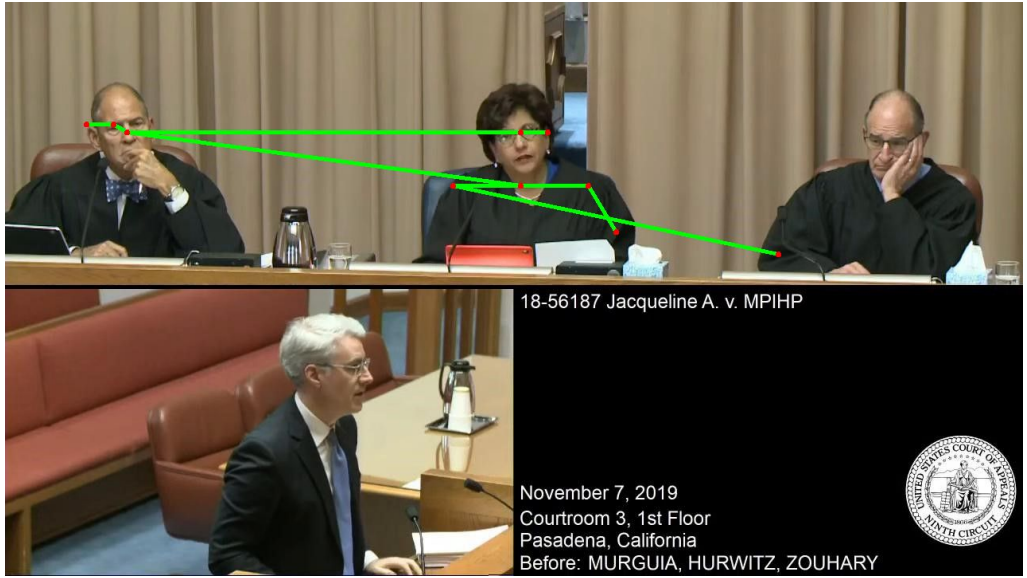
- Emotional Arousal in audio to predict Supreme Court votes (Dietrich, Enos, and Sen 2019)
- Audio analysis of telephone interviews (Dietrich, Mondak, Williams 2019)
- Measuring polarization with C-SPAN videos (Dietrich 2021)
- Analyzing political slant of images during movements like BLM (Kim and Bas 2023)

Judge Identification and Facial Emotion Recognition

- CVAT Labeling
  - Ninth Circuit Judges
  - Not labeled:
    - Visiting judges
    - Attorneys
    - Audience members/clerks/staff
- Facial Recognition
  - Two steps
  - Identify human faces and draw bounding boxes
  - Make prediction about who is in the bounding box (Haar Cascades, OpenFace Deep Neural Net, custom Convolutional Neural Net)
- Facial Emotion Recognition
  - One of seven emotions (happy, angry, disgusted, fearful, neutral, sad, surprised)
  - Record max emotion and emotion probability proportions

Judge Poses

- OpenPose framework
- Detect features corresponding to human body parts
- Construct features pairs to estimate “poses”
- To Do:
  - Label and classify specific poses
  - Match poses with outcomes like reading notes, looking at attorney/colleagues, etc.

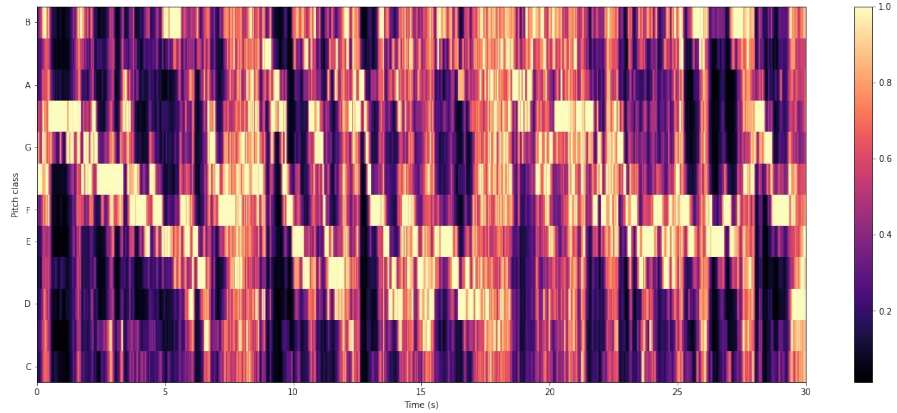


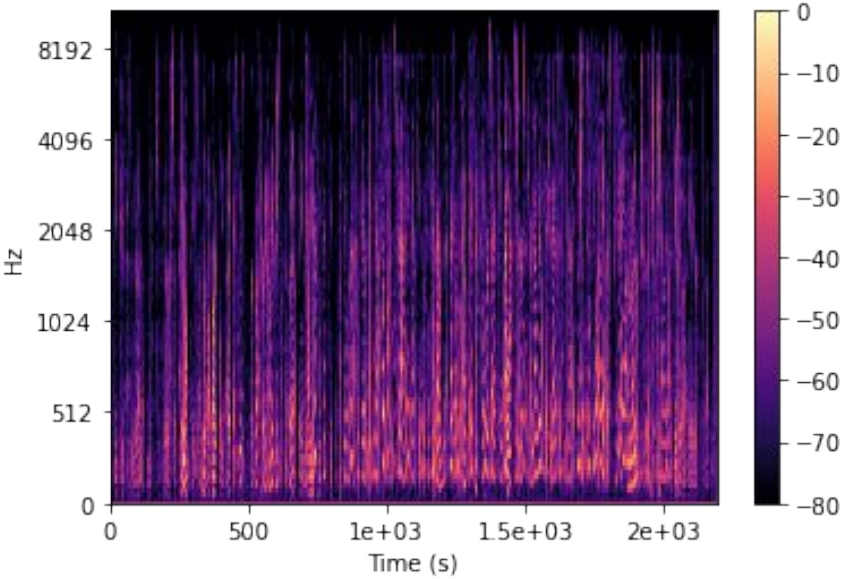
Speaker Diarization and Transcription

- Diarization and Transcription
  - OpenAI Whisper model for automatic diarization and transcription of text
    - Does NOT do speaker recognition
- Pipeline for segmenting by speaker and labeling transcription with our audio classification model

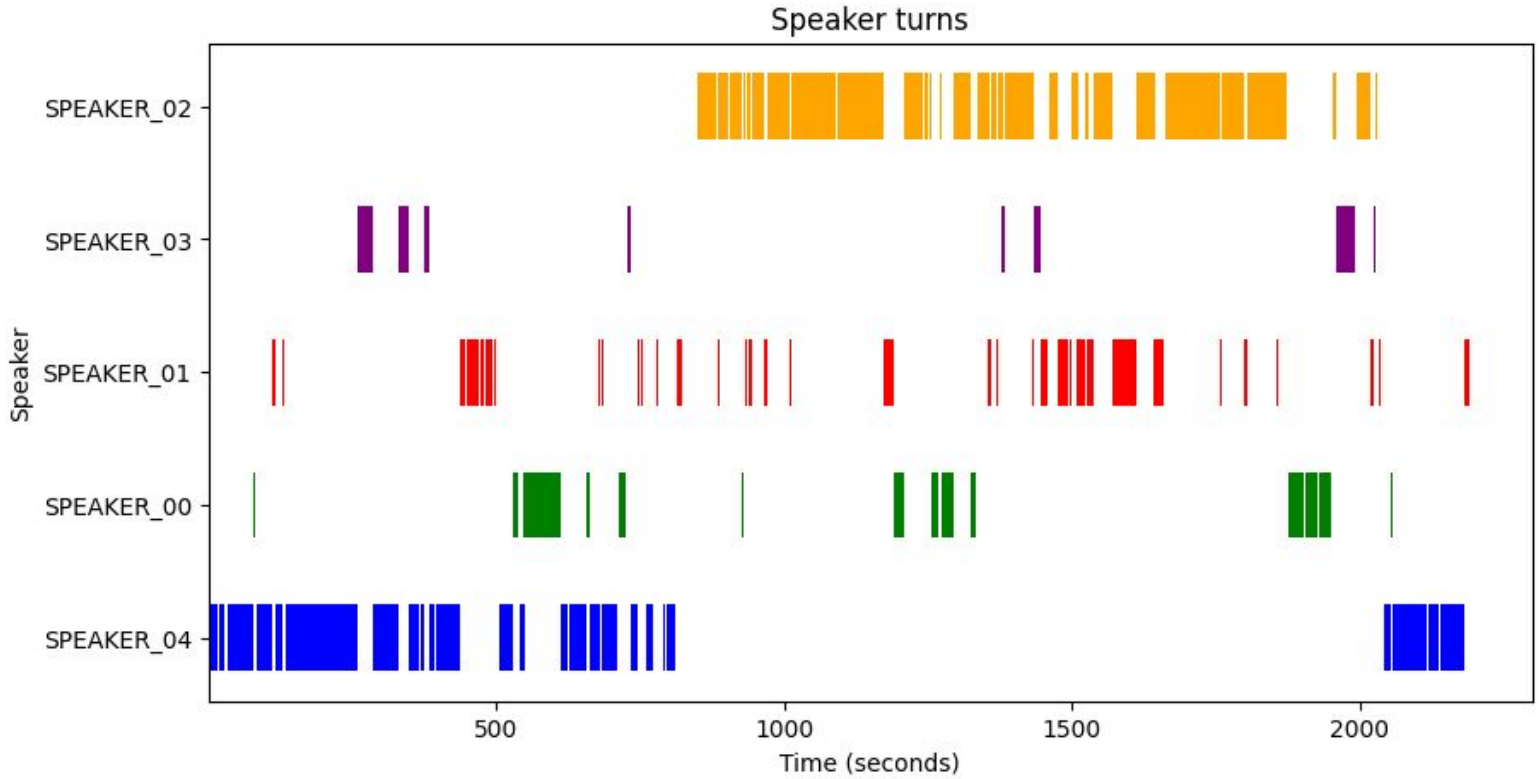
Audio Features

- Pitch, frequency, and amplitude
- Mel-frequency cepstral coefficients (MFCCs)
- Distinguish speakers
- “Emotional Arousal”

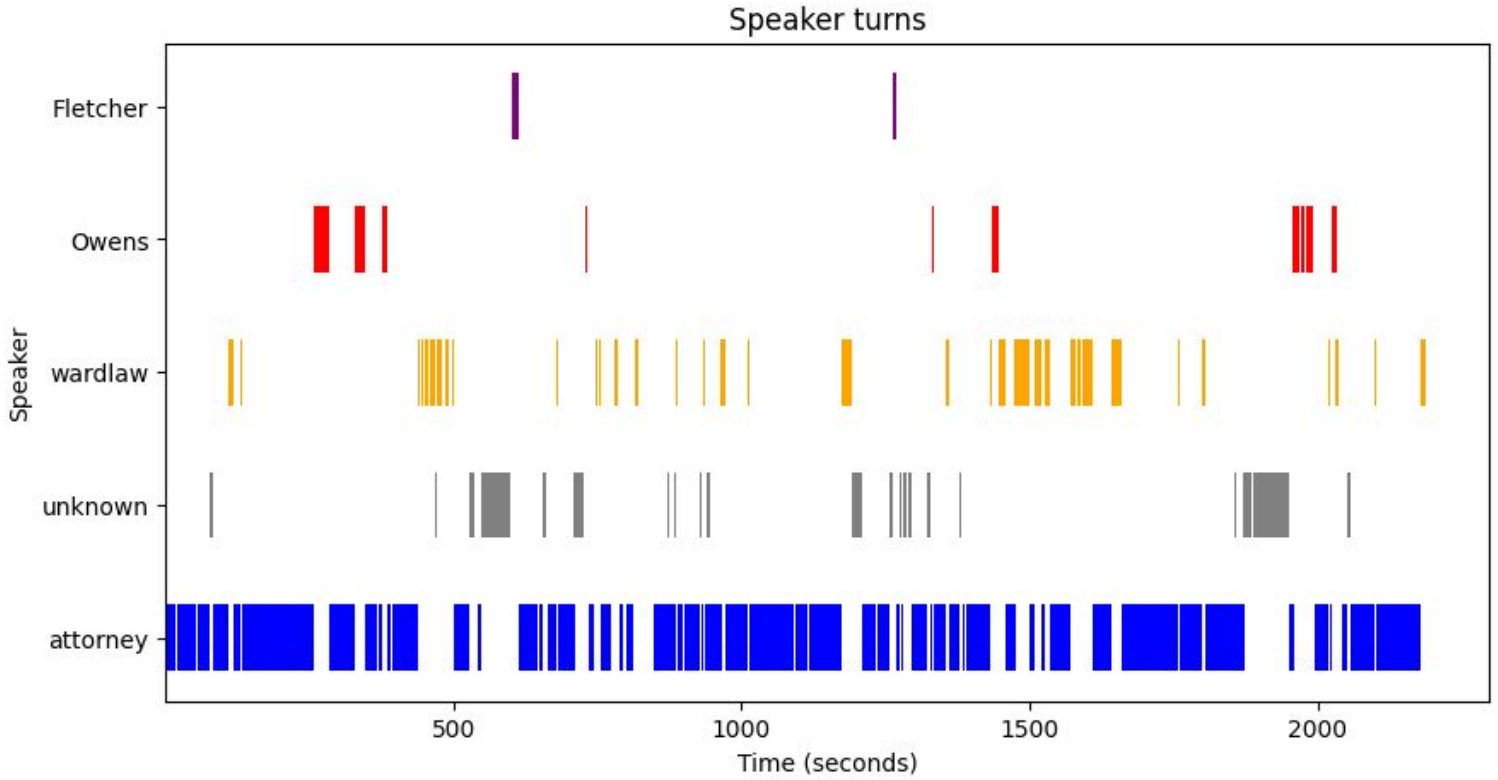




Diarization



Speaker Identification



Transcription

attorney: The expert testified for over an hour as to his qualifications. The experts testified that he had been a citizen in the United States since 2000., '

fletcher: Was there some reason why the expert didn't want to provide the A number? Was it hard to find?, '

attorney: The declaration of the expert states that he felt that he had been was being abused by the judge and that he did not want to provide any further testimony than he had., '

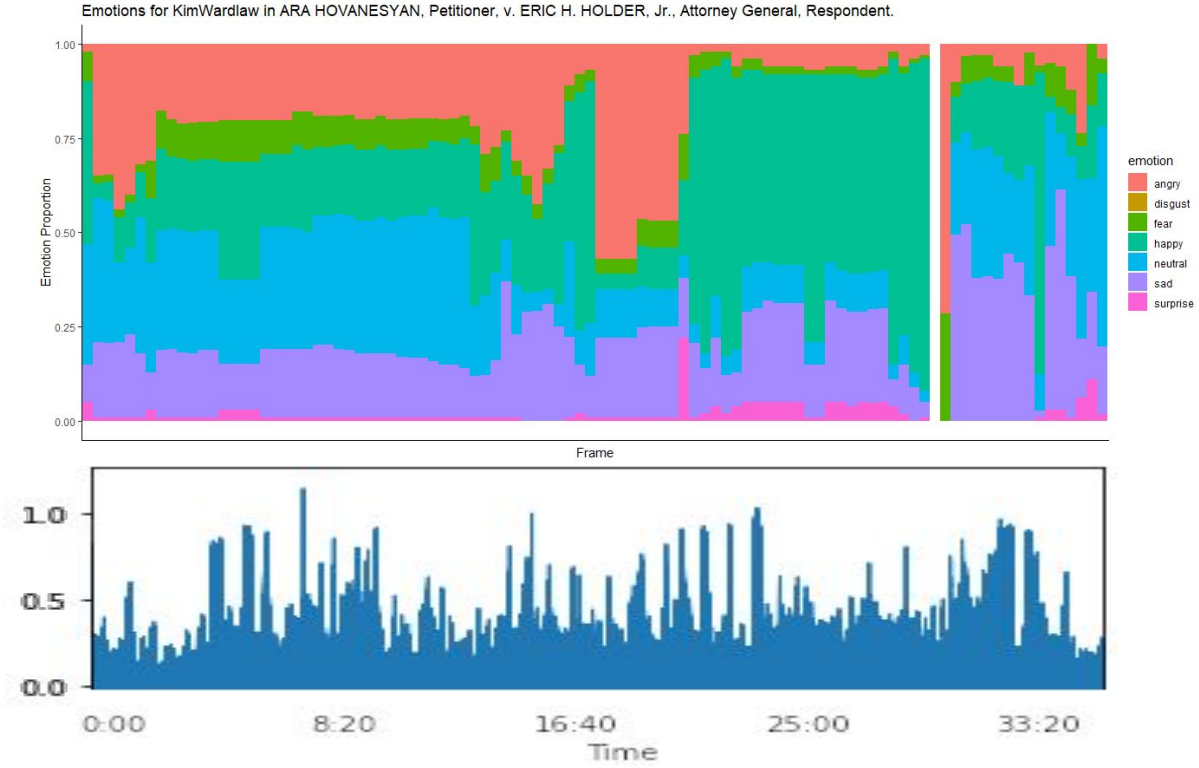
attorney: His declaration, the most emotional accuse, states that he was being unreasonably challenged by the judge, the judge is rolling her eyes., '

wardlaw: Is it an extended practice to ask for the A number of witnesses and immigration proceedings?, '

attorney: That would be true for a lay witness, Your Honor, but not necessarily for an expert witness. I don't see the relevance of providing an A number for the expert witnesses., '

wardlaw: Why are we assuming that the witness has an A number?, '

Tracking a Judge Through a Video



Tracking Judges Through a Video

Next Steps

- Labeling
- Automatically detecting key events and Higher Level Features
  - Judge making eye contact with colleagues and attorney, or looking at notes?
  - Interrupting attorneys and colleagues?
  - Relative to their own baselines, are judges bored? Sleepy? Engaged?
- Case-level analysis
  - Do judges behave differently in criminal and civil suits?
  - Do Republicans and Democrats behave differently?
  - How predictive are audio/video features of judge votes?
- Other future directions?
  - Applications of video + audio + generated text
  - Sentiment analysis
  - Audiovisual topic modeling in combination with text
  - Other areas?