


PROBLEM?

Humans are great liars but poor lie detectors.

(Ability of humans to detect deception without any special aids is around 54%[1])

V. Gupta, M. Agarwal, M. Arora, T. Chakraborty, R. Singh and M. Vatsa, "Bag-of-Lies: A Multimodal Dataset for Deception Detection," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 2019, pp. 83-90, doi: 10.1109/CVPRW.2019.00016.

Polygraphs?
NO.
but WHY?

- Requirement of specialized personnel
- Invasive by nature[2]
- Not Scalable (expensive and human presence required)
- Ethical Concerns

[2] Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing: Scientific analysis and public controversy. American Psychologist, 40(3), 355.

DECEPTION DETECTION

using non-Intrusive Methods

Chaitanya Modi, Hibah Ihsan Muhammad, Subham Jalan

Healthcare

Verify patient testimony

Corporate

Verify credentials

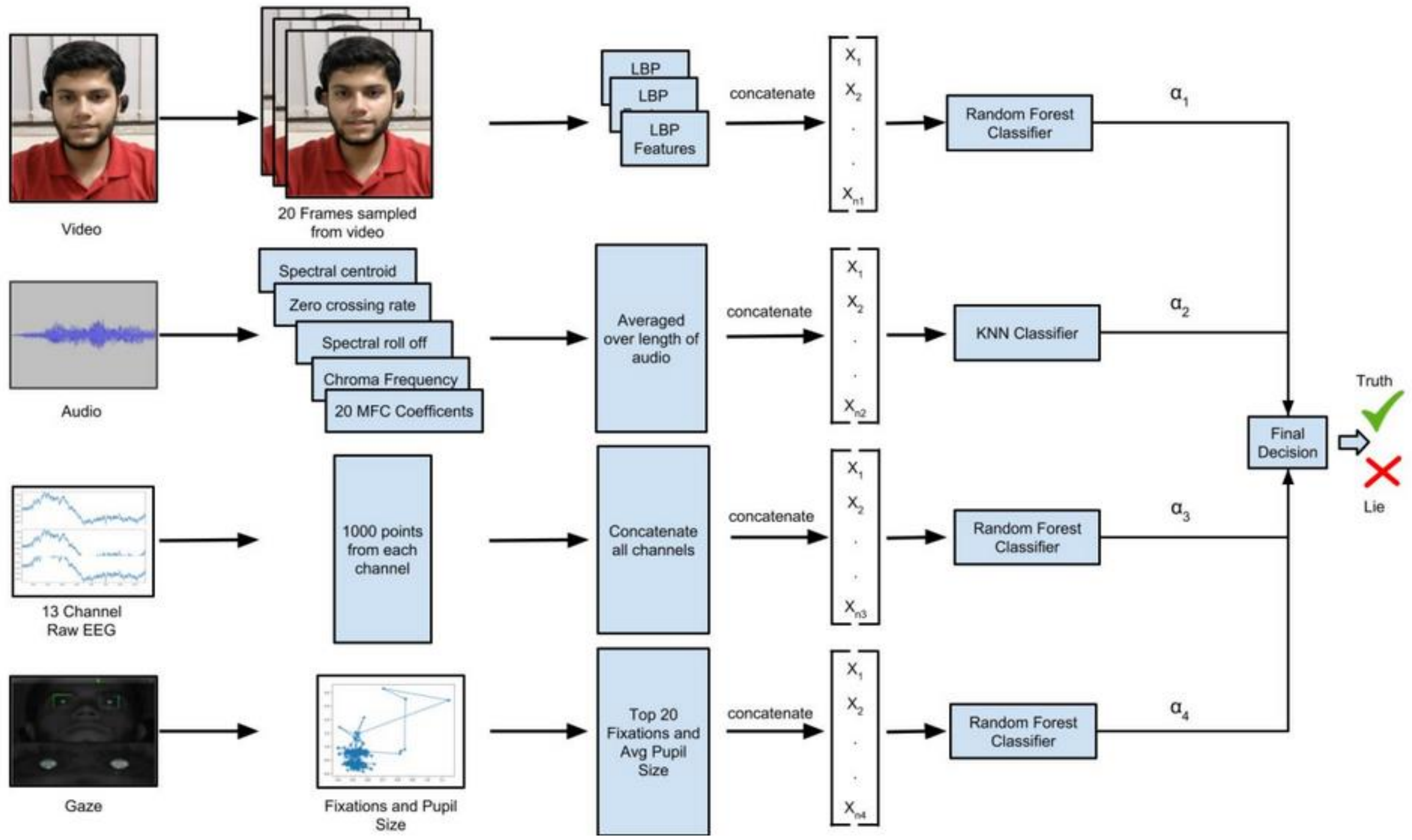
Education

Academic Dishonesty

Literature Survey

Paper 1

Bag-of-Lies: A Multimodal Dataset for Deception Detection[1]



[1] V. Gupta, M. Agarwal, M. Arora, T. Chakraborty, R. Singh and M. Vatsa, "Bag-of-Lies: A Multimodal Dataset for Deception Detection," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 2019, pp. 83-90, doi: 10.1109/CVPRW.2019.00016.

Methodology:

Proposed a late fusion model combining predictions from individual models working on different modalities

Observations:

- Limited set of features (e.g. only LBP for video)
- Weighted late fusion (11-fold cross validation)

Performance Metric:

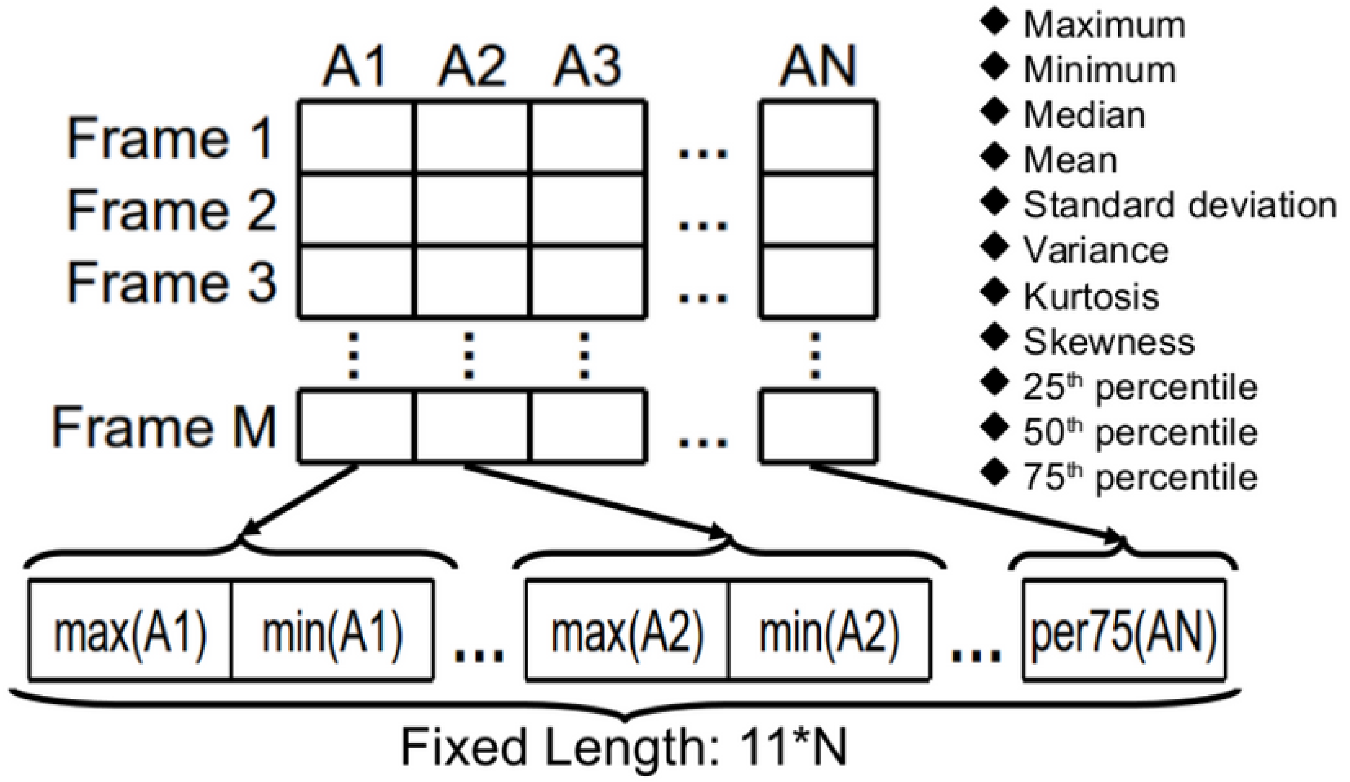
Accuracy (all modalities): 66.17%

Analysis:

- Gaze can be extracted from video
- Neural Networks can be tried for higher accuracy

Paper 2

High-level Features for Multimodal Deception Detection in Videos[2]



Methodology:

Studies different high-level features that can be extracted from video, audio, and text using open-source tool and their fusion techniques.

Observations:

- Late fusion outperformed early fusion (0.05AUC)
- Hyperparameter tuning not done

Performance Metric:

AUC (all modalities): 67%

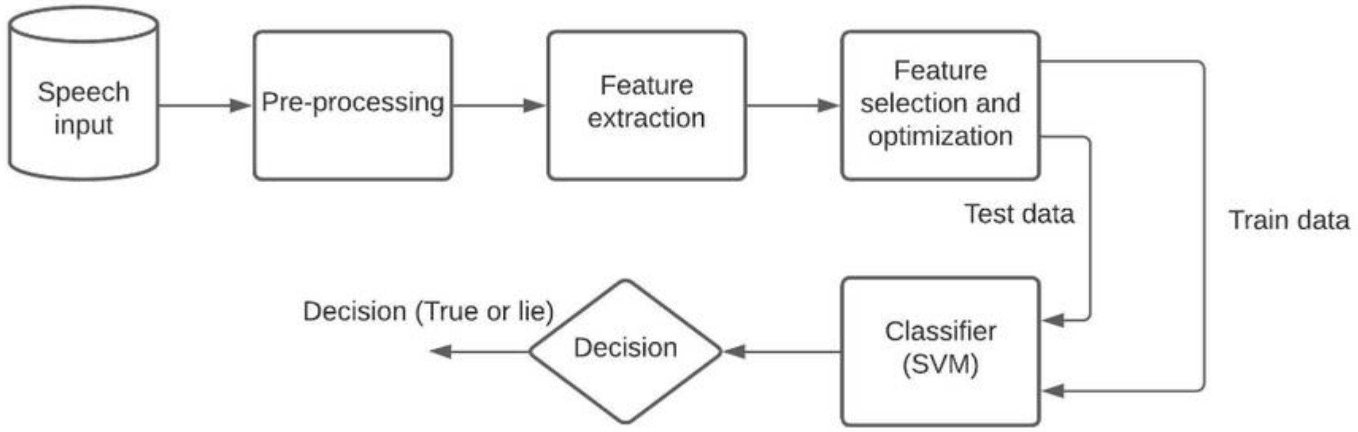
Analysis:

- Variable-length sequences can be converted to statistical embeddings
- OpenFace (library) can be used to extract more video features.
- Gaze is backed by research for deception detection. (>0.5AUC)

[2] R. Rill-García, H. J. Escalante, L. Villaseñor-Pineda and V. Reyes-Meza, "High-Level Features for Multimodal Deception Detection in Videos," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 2019, pp. 1565-1573, doi: 10.1109/CVPRW.2019.00198.

Paper 3

Lie Detection using Speech Processing Techniques[3]



[3] E. P. Fathima Bareeda, S. Mohan, and K. V. Ahammed Muneer, "Lie Detection using Speech Processing Techniques," Journal of physics. Conference series, vol. 1921, no. 1, pp. 012028–012028, May 2021, doi: https://doi.org/10.1088/1742-6596/1921/1/012028.

Methodology:

Uses MFCC from speech signals and SVM classifier for lie detection in isolated speech utterances.

Observations:

- Only MFCCs were extracted
- Small dataset (161) may limit model generalization.

Performance Metric:

Accuracy (Audio): 81.48%

Analysis:

- Dimensionality reduction can be done using PCA to avoid overfitting.
- More features could be extracted.

Paper 4

UNIVERSITY OF MICHIGAN

Multimodal Deception Detection Using Real-Life Trial Data[4]

Individual Feature Performance: Accuracy (%) and AUC Scores

Feature Set (dimension)	SVM		RF		NN	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Visual						
Facial displays (32)	76.27± 0.00	0.8581	76.27 ± 1.69	0.9270	**80.79** ± 0.98	0.9416
Hand gestures (7)	50.28 ± 3.53	0.7232	**64.97** ± 3.91	0.6671	61.58 ± 0.98	0.6930
All visual (39)	58.19 ± 0.98	0.8641	77.40 ± 0.98	0.9187	**78.53** ± 1.96	0.9377
Acoustic						
Pitch (std-f_0) (1)	61.58 ± 0.98	0.6507	**71.19** ± 3.39	0.7939	51.41 ± 0.98	0.7427
Pitch (mean-f_0) (1)	54.24 ± 1.69	0.5223	53.11 ± 0.98	0.5465	**61.02** ± 0.00	0.5235
Sil.Sp.Hist (50)	57.63 ± 0.00	0.4159	**59.32** ± 2.94	0.7069	55.93 ± 1.69	0.6483
All Acoustic (52)	56.50 ± 2.59	0.5864	**63.28** ± 0.98	0.7059	61.02 ± 4.48	0.6589
Linguistic						
Unigrams (134)	53.11 ± 1.96	0.7275	**64.41** ± 4.48	0.6173	63.28 ± 0.98	0.7651
Unigrams - LIWC (100)	52.54 ± 4.48	0.5906	**63.84** ± 2.59	0.6764	55.93 ± 1.69	0.7729
All Linguistic (234)	53.11 ± 4.27	0.6765	**61.58** ± 2.59	0.6605	57.63 ± 1.69	0.7655

[4] U. M. Sen, V. Perez-Rosas, B. Yanikoglu, M. Abouelenien, M. Burzo and R. Mihalcea, "Multimodal Deception Detection using Real-Life Trial Data," in IEEE Transactions on Affective Computing, vol. 13, no. 1, pp. 306-319, 2022.

Methodology:

Uses lingiustic, visual and acoustic features on real life deception detection by fusing the three modalities.

Observations:

- Meant for High Stake deception detection
- Neural network worked the best with 8055 number of words.

Performance Metric:

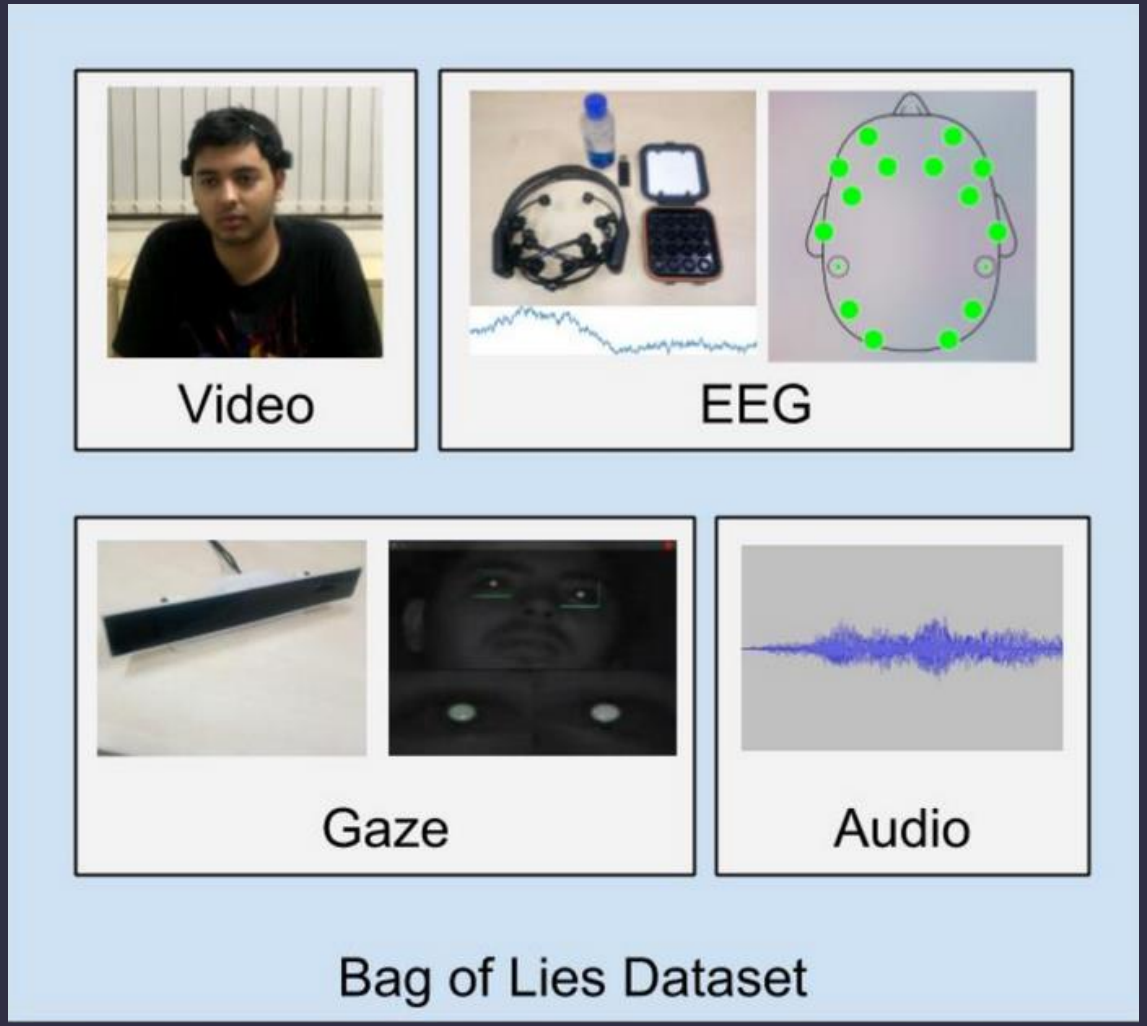
Accuracy (all modalities): 84.18%

Analysis:

- RF and SVM can be used for transcript classification.
- Unigrams can be extracted for transcript modality

Dataset

- Procured from **IIIT-Delhi**[5]
- Modalities - Video, EEG, Gaze, Audio
- Specialized equipment
  (Emotiv EPOC+EEG, Gazepoint GP3)



[5] V. Gupta, M. Agarwal, M. Arora, T. Chakraborty, R. Singh and M. Vatsa, "Bag-of-Lies: A Multimodal Dataset for Deception Detection," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 2019, pp. 83-90, doi: 10.1109/CVPRW.2019.00016.

Dataset

Experimental Setup:

- Participants shown 6-10 images
- Asked to describe final image
- **Natural** choice of deception!

Provides a non-hypothetical deception scenario for casual deceptions

Sample Images

Photo Masked

Dataset

Properties:

- 325 samples - 162 (L) + 163 (T)
- 10 females and 25 males
- Model performed 2.3% better on males
- Thick hair obstruction

Ethical Considerations:

- Restricted Dataset
- Secure server - access control

Data Preprocessing

Transcripts

Step 1: Extraction of Transcript from Audio

Original Number of Data points: 325

↓ Whisper v3 OpenAI

Number of Data Points after extracting Transcript

Step 2: Creation of a Dataframe

Original Dataframe

Data Masked

Preprocessed Dataframe

Data Masked

Basic Preprocessing

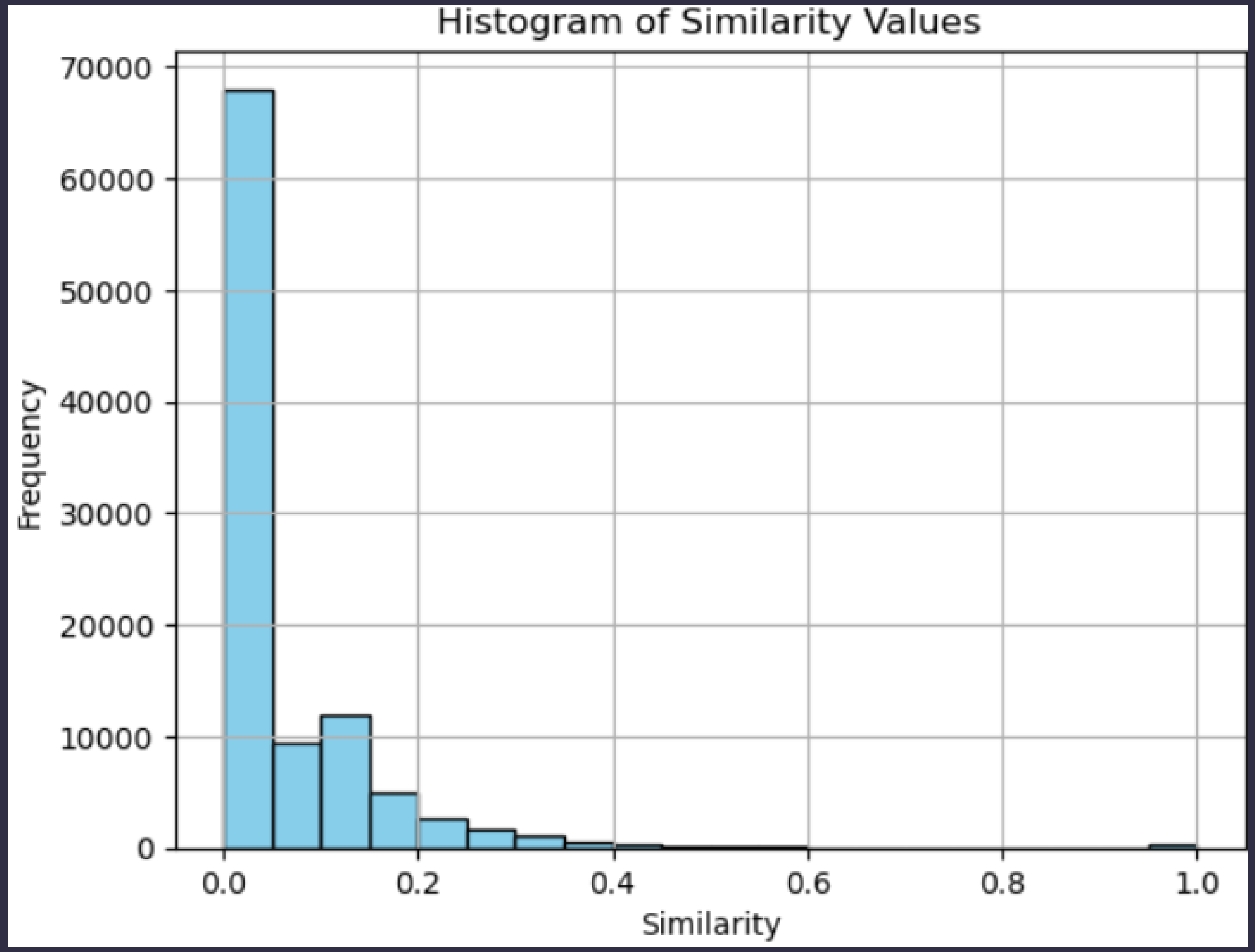
- Lowercasing
- Removing Punctuation
- Removing Stopwords
- Handling Special Characters
- **Removing HTML Tags and URLs**

Transcripts

Step 3: Make a Cosine similarity Matrix and Plot a histogram

Similarity Matrix

```
[[1.         0.         0.1490712  ... 0.         0.         0.11111111]
 [0.         1.         0.         ... 0.14285714 0.         0.        ]
 [0.1490712  0.         1.         ... 0.         0.         0.        ]
 ...
 [0.         0.14285714 0.         ... 1.         0.         0.12598816]
 [0.         0.         0.         ... 0.         1.         0.        ]
 [0.11111111 0.         0.         ... 0.12598816 0.         1.        ]]
```



Step 4: Decide a Threshold according to the Histogram and filter the df

```
result
0    124
1     81
```

Final Filtered Dataframe shape: (205,1)

Transcripts

Step 5: Feature Extraction

Bag of Words(BOW)

```
Bag of Words (BoW):
 [[0 0 0 0 0 0 0 1 0 1 0 0 0 1 0 1 0 1 0 1 2 0 0 0 1 0 0 0]
 [1 0 1 1 0 1 0 0 0 0 1 1 1 0 1 0 0 0 1 0 1 1 0 0 1 1 1 1]
 [0 1 0 0 1 0 1 0 1 0 0 0 0 0 0 0 1 0 1 0 1 0 1 1 0 0 0 0]]
```

N-grams

```
N-grams:
 [[0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1 0 0 0 0 0 1 1 0 0 1 0 0 1 1 0 0
  0 1 1 2 0 1 0 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0]
 [1 1 0 0 1 1 1 1 0 0 1 1 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 1 1 0 0 0 0 0 1 0
  1 0 0 1 0 0 1 0 1 1 0 0 0 0 1 0 1 1 1 1 1 1]
 [0 0 1 1 0 0 0 0 1 1 0 0 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1
  0 0 0 1 1 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0]]
```

Transcripts

TF-IDF

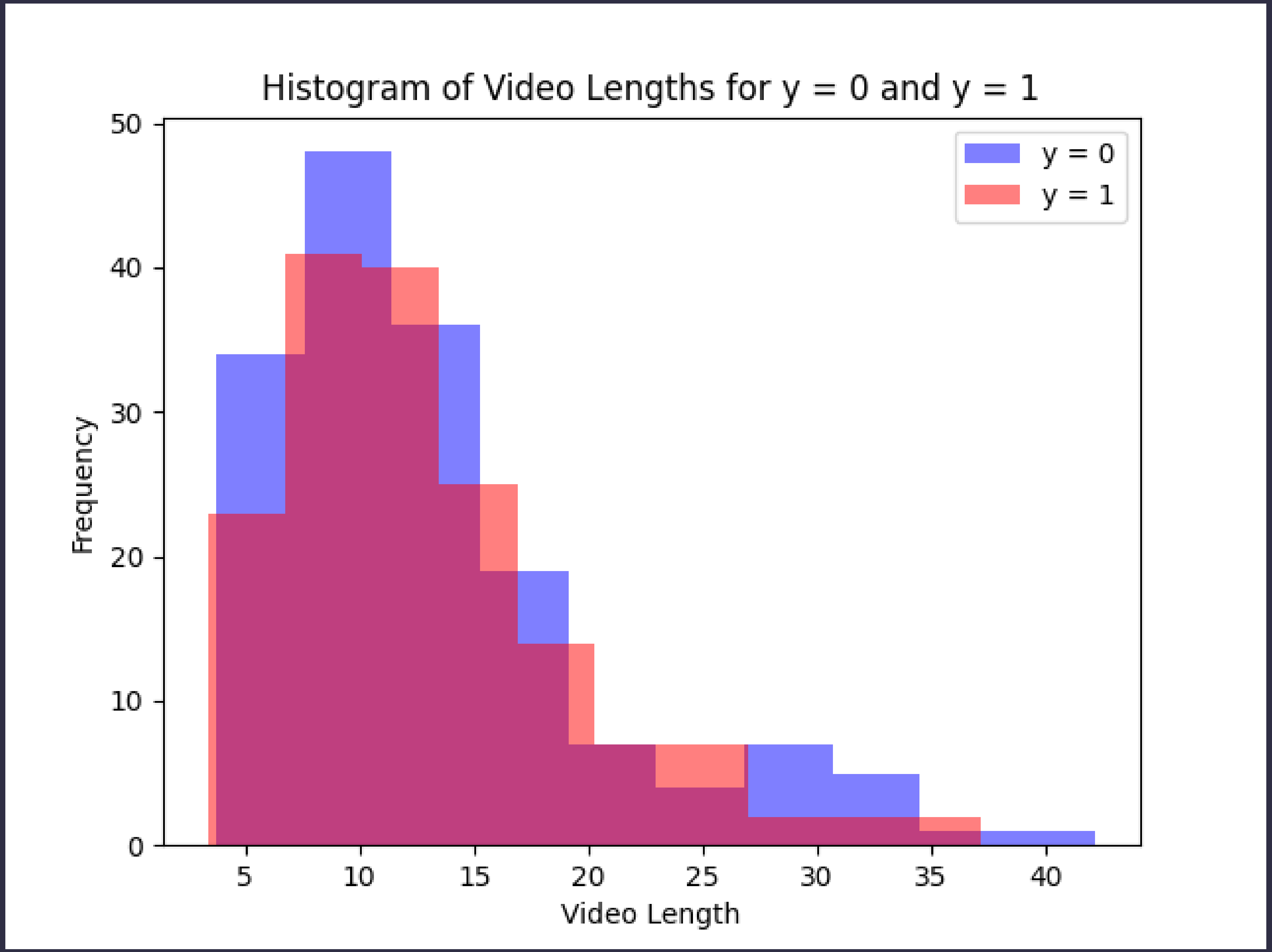
```
TF-IDF:
 [[0.         0.         0.         0.         0.         0.
  0.         0.35413578 0.         0.35413578 0.         0.
  0.         0.35413578 0.         0.35413578 0.         0.35413578
  0.         0.35413578 0.41831659 0.         0.         0.
  0.26932939 0.         0.         0.        ]
 [0.27210883 0.         0.27210883 0.27210883 0.         0.27210883
  0.         0.         0.         0.         0.27210883 0.27210883
  0.27210883 0.         0.27210883 0.         0.         0.
  0.20694578 0.         0.16071186 0.27210883 0.         0.
  0.20694578 0.27210883 0.27210883 0.27210883]
 [0.         0.35517252 0.         0.         0.35517252 0.
  0.35517252 0.         0.35517252 0.         0.         0.
  0.         0.         0.         0.         0.35517252 0.
  0.27011786 0.         0.20977061 0.         0.35517252 0.35517252
  0.         0.         0.         0.        ]]
```

Word2Vec

```
Word2Vec:
[ 8.3474135e-03 -5.7207549e-04 -9.4375769e-03  4.7827936e-03
 -6.0445508e-03  6.6838926e-03  5.3725867e-03 -5.0468231e-03
  2.5687546e-03  5.4189041e-03 -3.5865146e-03 -1.5144218e-03
  9.1719013e-03  9.0625333e-03 -9.3910722e-03  7.5650970e-03
  9.8875528e-03 -2.8391804e-03  2.4573463e-03 -2.8026837e-03
  8.6430376e-03 -2.8467341e-04  5.6364359e-03  9.2138294e-03
  4.1098148e-03 -7.1207187e-03 -1.9234006e-03  9.7789941e-04
  2.0326125e-03  2.9556018e-03  9.4502280e-03  4.4002603e-03
  9.9119144e-03 -8.6575756e-03 -5.7540201e-03  1.9874072e-03
  3.6548518e-03 -9.9221768e-04 -6.9129714e-03 -3.2113763e-03
 -8.5284319e-03  9.4111022e-03  3.7243692e-03 -7.8788400e-03
  3.1879896e-03  4.1732127e-03 -5.6372448e-03 -5.9139454e-03
  1.0364689e-03  8.9602843e-03 -9.6455161e-03  5.6382296e-06
 -6.8661813e-03 -9.3362684e-04  3.0388411e-03 -5.0303270e-03
 -2.7774265e-03  6.7519158e-04 -6.3632787e-03  7.2843963e-03
  4.3802024e-03 -8.5593462e-03 -2.1482927e-03  3.1643168e-03
 -8.3279693e-03 -7.0694438e-03 -8.4527740e-03 -5.4971008e-03
  8.8549480e-03  7.0773163e-03  2.8861596e-03 -8.5535981e-03
```

Video

Step 1: Data Familiarization




```
Minimum Length: 3.3621777777777777
Maximum Length: 42.166666666666664
Average Length: 13.048290940566174
```

Video

Step 2: Extract Features (OpenFace)

Photo Masked

Extracted features overlay

frame	face_id	timestamp	confidence	success	gaze_0_x	gaze_0_y	gaze_0_z	gaze_1_x	gaze_1_y	gaze_1_z	gaze_angle_x
1	0	0.000	0.98	1	0.156237	0.498489	-0.852701	-0.155730	0.515231	-0.842784	0.000	0.539	259.6
2	0	0.033	0.98	1	0.146927	0.497123	-0.855150	-0.158016	0.487946	-0.858452	-0.006	0.522	258.9
3	0	0.067	0.98	1	0.154291	0.485324	-0.860613	-0.164344	0.477369	-0.863198	-0.006	0.509	259.0
4	0	0.100	0.98	1	0.152318	0.481375	-0.863179	-0.173090	0.476330	-0.862061	-0.012	0.507	259.1
5	0	0.133	0.98	1	0.153508	0.483944	-0.861530	-0.168420	0.474387	-0.864055	-0.009	0.507	258.9
6	0	0.166	0.98	1	0.167302	0.487480	-0.856956	-0.166816	0.482054	-0.860114	0.000	0.514	259.0
7	0	0.200	0.98	1	0.172715	0.471560	-0.864755	-0.166749	0.485960	-0.857927	0.003	0.507	258.9
8	0	0.233	0.98	1	0.158576	0.474541	-0.865832	-0.146456	0.484456	-0.862469	0.007	0.507	258.0
9	0	0.266	0.98	1	0.164370	0.481046	-0.861149	-0.159547	0.486407	-0.859042	0.003	0.512	258.0
10	0	0.300	0.98	1	0.170122	0.475289	-0.863226	-0.154423	0.477503	-0.864953	0.009	0.504	257.8
11	0	0.333	0.98	1	0.170398	0.473574	-0.864114	-0.140283	0.483219	-0.864187	0.017	0.506	257.7
12	0	0.366	0.98	1	0.171272	0.468305	-0.866808	-0.161096	0.471278	-0.867148	0.006	0.497	257.2
13	0	0.399	0.98	1	0.171319	0.456795	-0.872919	-0.161271	0.467526	-0.869144	0.006	0.488	256.7
14	0	0.433	0.98	1	0.171955	0.472245	-0.864532	-0.142391	0.482240	-0.864389	0.017	0.504	256.4
15	0	0.466	0.98	1	0.179500	0.474240	-0.861903	-0.146723	0.477667	-0.866203	0.019	0.503	256.2
16	0	0.499	0.98	1	0.183287	0.473235	-0.861658	-0.142830	0.482837	-0.863984	0.023	0.506	256.1
17	0	0.533	0.98	1	0.178140	0.474450	-0.862069	-0.133560	0.488981	-0.862009	0.026	0.510	255.7
18	0	0.566	0.98	1	0.174425	0.478418	-0.860635	-0.139020	0.488504	-0.861416	0.021	0.512	255.3
19	0	0.599	0.98	1	0.171640	0.476181	-0.862433	-0.155734	0.476381	-0.865337	0.009	0.504	255.1
20	0	0.632	0.98	1	0.175253	0.477866	-0.860773	-0.154496	0.483662	-0.861512	0.012	0.509	255.0
21	0	0.666	0.98	1	0.180504	0.477152	-0.860084	-0.152572	0.479457	-0.864200	0.016	0.507	254.8
22	0	0.699	0.98	1	0.175610	0.479423	-0.859834	-0.138994	0.489909	-0.860622	0.021	0.513	254.9
23	0	0.732	0.98	1	0.176967	0.472934	-0.863143	-0.153054	0.478969	-0.864386	0.014	0.504	255.0
24	0	0.766	0.98	1	0.177460	0.481628	-0.858220	-0.149103	0.498052	-0.854232	0.017	0.520	255.2
25	0	0.799	0.98	1	0.170422	0.487813	-0.856151	-0.154129	0.493881	-0.855760	0.010	0.521	255.5
26	0	0.832	0.98	1	0.172349	0.490304	-0.854341	-0.152962	0.490870	-0.857700	0.011	0.520	256.2
27	0	0.866	0.98	1	0.169141	0.485012	-0.857995	-0.144625	0.494994	-0.856776	0.014	0.519	256.7

68 Facial Landmarks, 18 Action Units, 2 eyes' Gaze, 56 eye landmarks

Video

Other Features

Photo Masked

478 Mediapipe Landmarks

Photo Masked

Surprise

Photo Masked

Disgust

DeepLie Facial Expressions (total 7)

Video

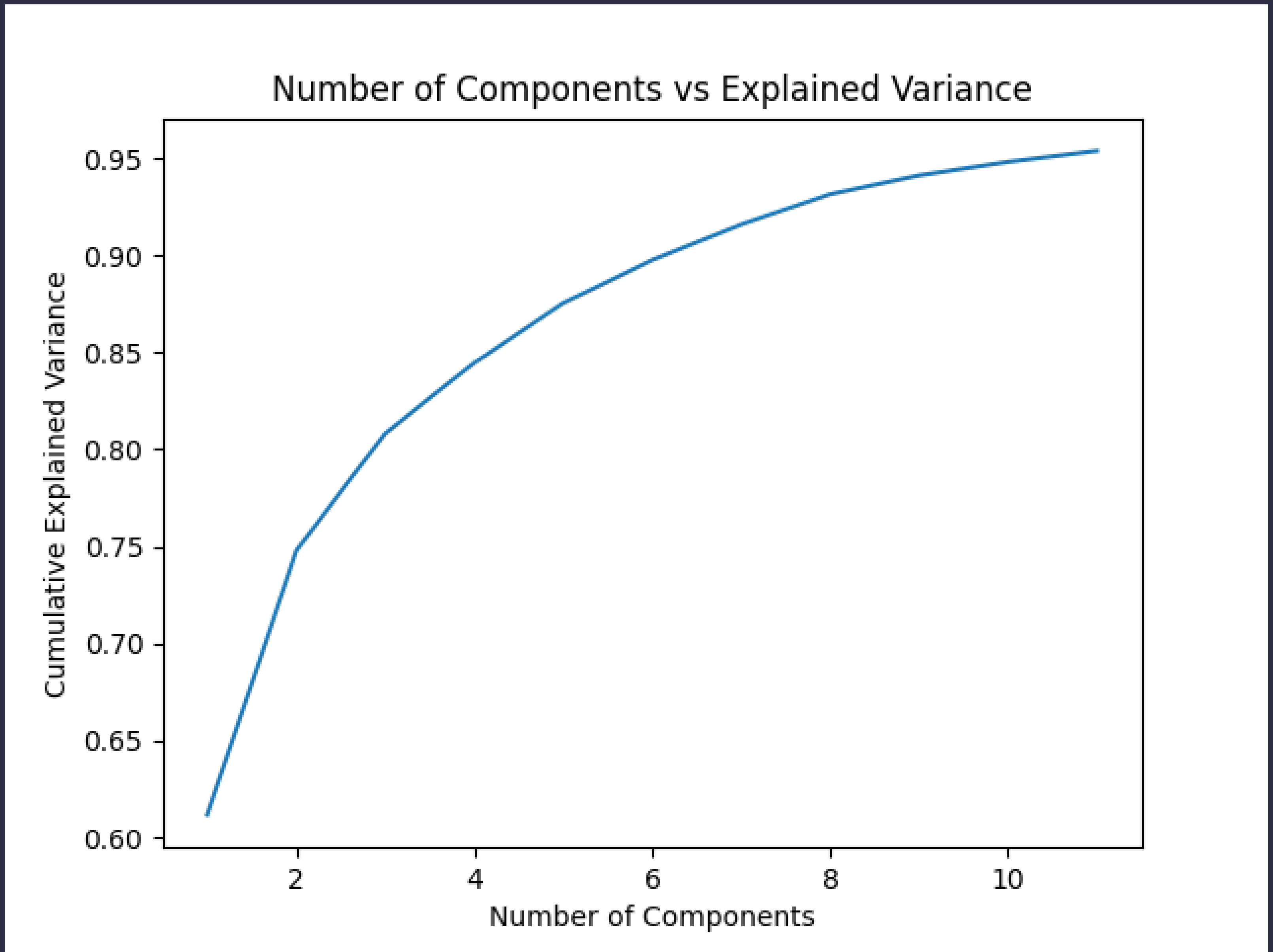
Step 3: Convert to Statistical Embeddings

	0	1	2	3	4	5	6	...	1427	1428	1429	1430	1431	1432	1433
0	0.628220	0.585478	-0.021724	0.617518	0.572269	-0.042757	0.621676	...	0.001981	0.640676	0.541239	0.001970	0.650270	0.543611	0.001973
1	0.628009	0.585624	-0.021663	0.617510	0.572577	-0.042757	0.621563	...	0.001979	0.640296	0.542018	0.001968	0.649334	0.544293	0.001970
2	0.627554	0.585871	-0.021660	0.617374	0.572803	-0.042769	0.621241	...	0.002123	0.641028	0.542481	0.002113	0.649646	0.544645	0.002115
3	0.626945	0.586212	-0.021498	0.617005	0.572950	-0.042704	0.620718	...	0.002077	0.639203	0.542973	0.002067	0.647669	0.545110	0.002069
4	0.626104	0.586403	-0.021498	0.616056	0.573038	-0.042702	0.619783	...	0.001727	0.637685	0.543259	0.001717	0.646416	0.545432	0.001719
..	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
521	0.625486	0.590213	-0.021346	0.615473	0.577408	-0.042215	0.619461	...	0.000501	0.638121	0.545078	0.000491	0.646948	0.547548	0.000492
522	0.625438	0.590783	-0.020850	0.615435	0.578410	-0.042217	0.619446	...	-0.000092	0.637960	0.545139	-0.000101	0.646833	0.547624	-0.000100
523	0.625128	0.591031	-0.020519	0.615129	0.578717	-0.042238	0.619133	...	-0.000579	0.637564	0.545143	-0.000588	0.646569	0.547629	-0.000587
524	0.624463	0.591075	-0.020342	0.614362	0.578790	-0.042243	0.618375	...	-0.000791	0.636464	0.545124	-0.000800	0.645539	0.547611	-0.000799
525	0.623673	0.590776	-0.020348	0.613669	0.578522	-0.042245	0.617681	...	-0.000787	0.636197	0.544619	-0.000797	0.645292	0.547097	-0.000795

136 x 11 = 1496 features

Video

Step 4: Apply PCA



	0	1	2	3	4	5	6	7	8	9	10
0	-33.480877	-5.368075	6.939137	-2.811792	-6.280956	7.498868	1.223844	3.118128	4.616891	1.369401	-0.946723
1	-34.385876	-5.591050	4.876374	0.196208	-6.636379	4.670253	2.048193	-6.568414	4.869732	0.413672	-2.675267
2	-30.210544	-5.123522	1.580824	-1.249155	-1.099228	1.396236	1.004438	-1.945191	2.348769	-1.418723	-1.690945
3	-34.382739	-2.224694	7.015470	-1.340062	-2.178119	10.978499	-1.833982	-4.183350	5.659691	-0.803370	-3.424389
4	-34.569953	-0.570395	12.460204	-2.954755	9.110812	18.752028	-10.461954	-2.750856	-0.138756	0.911343	-1.352495
..	...	...	...	...	...	...	...	...	...	...	...
320	-14.499753	1.462785	0.233516	-5.151235	-6.260486	-4.381239	0.449432	6.791107	3.356487	-4.297751	-1.554691
321	-12.714052	-4.377924	-1.696009	-4.239537	2.676205	-3.348984	-0.748482	-1.524912	-0.148899	-1.866274	-1.506226
322	-13.364502	-2.867546	2.690253	2.818882	1.476687	-1.468402	-5.403065	3.332910	0.636667	-1.608022	0.660161
323	-13.461854	-0.713506	1.260010	-4.369032	-0.592522	-2.460078	-3.791880	4.430088	-0.997241	-3.354472	0.332586
324	-13.025185	-2.475010	2.835088	-4.447439	7.458357	2.222256	-8.135903	-0.719897	-0.784331	-1.858201	0.481672

[325 rows x 11 columns]

Down to 11 features!

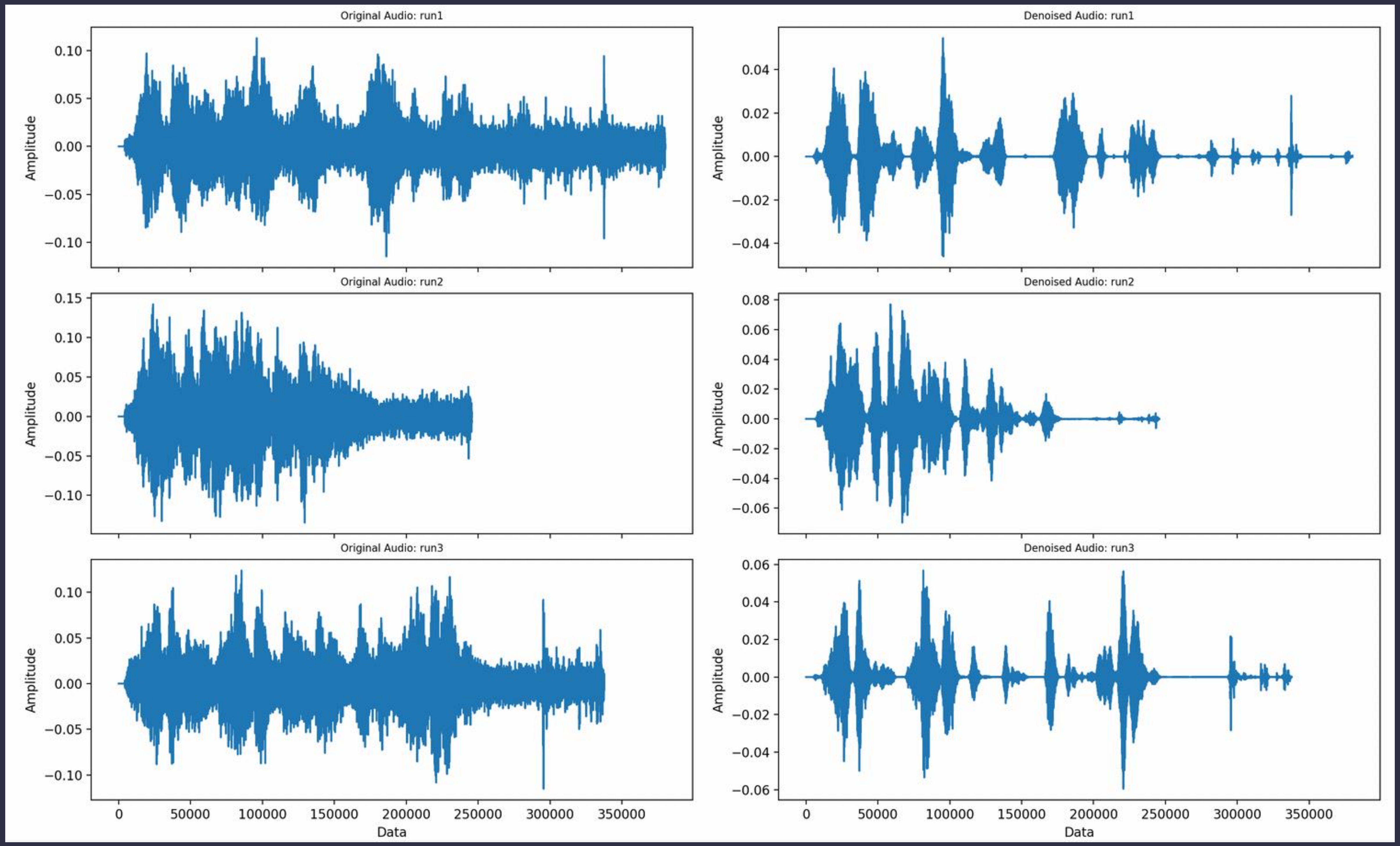
Audio

Step 1: Extract audio from video

- Using moviepy.editor
- Saving as wav files

Step 2: Extract denoised audio

- Using noisereduce
- Saving as new wav files

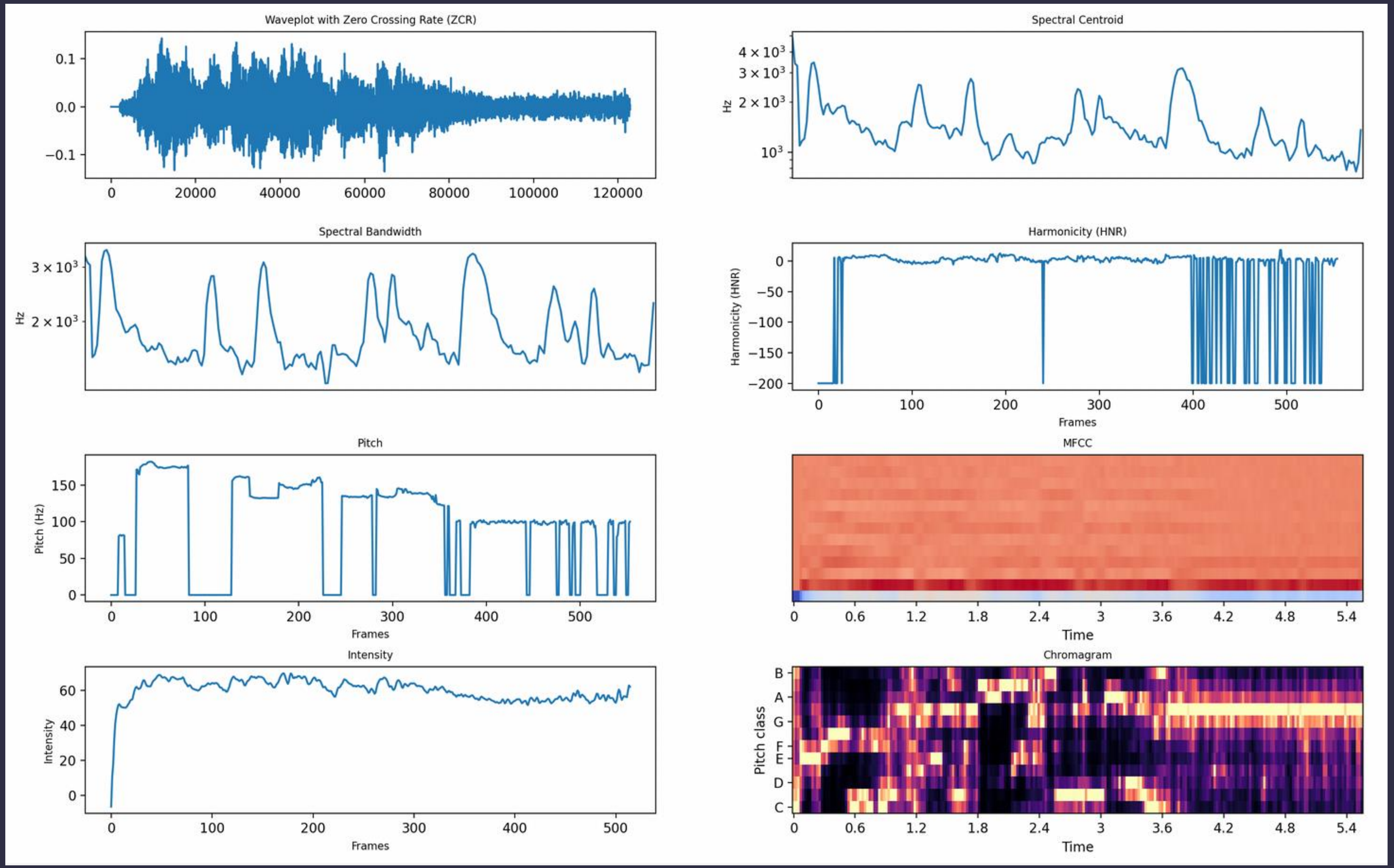


Noise and Denoised audio waveforms of a User

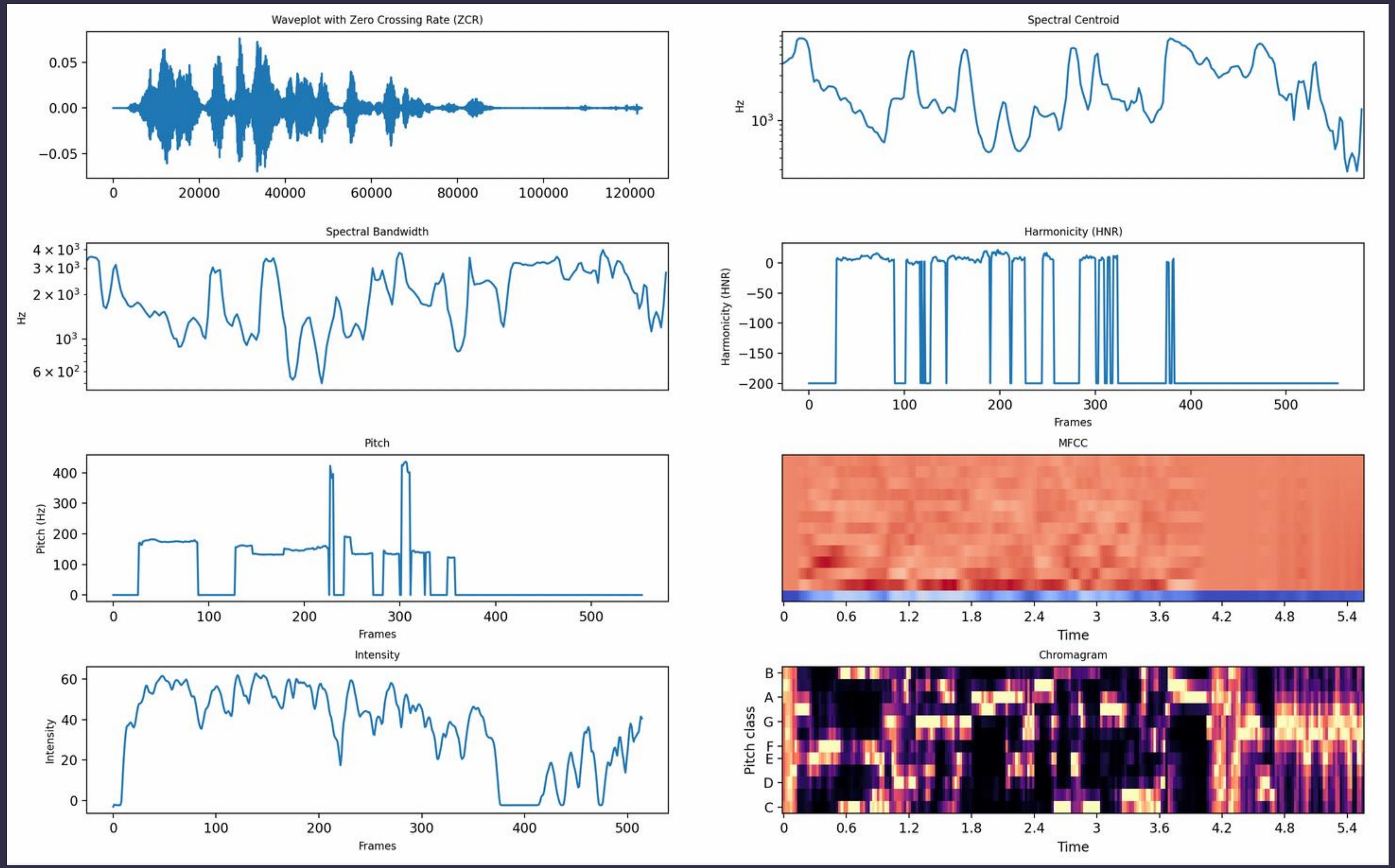
Audio

Step 3: Extract audio features

- Using librosa and parselmouth
- Total 18 features of varying dimensions
- Get the mean value for each dimension
- Normalize and save in a df



Features for Noise waveforms

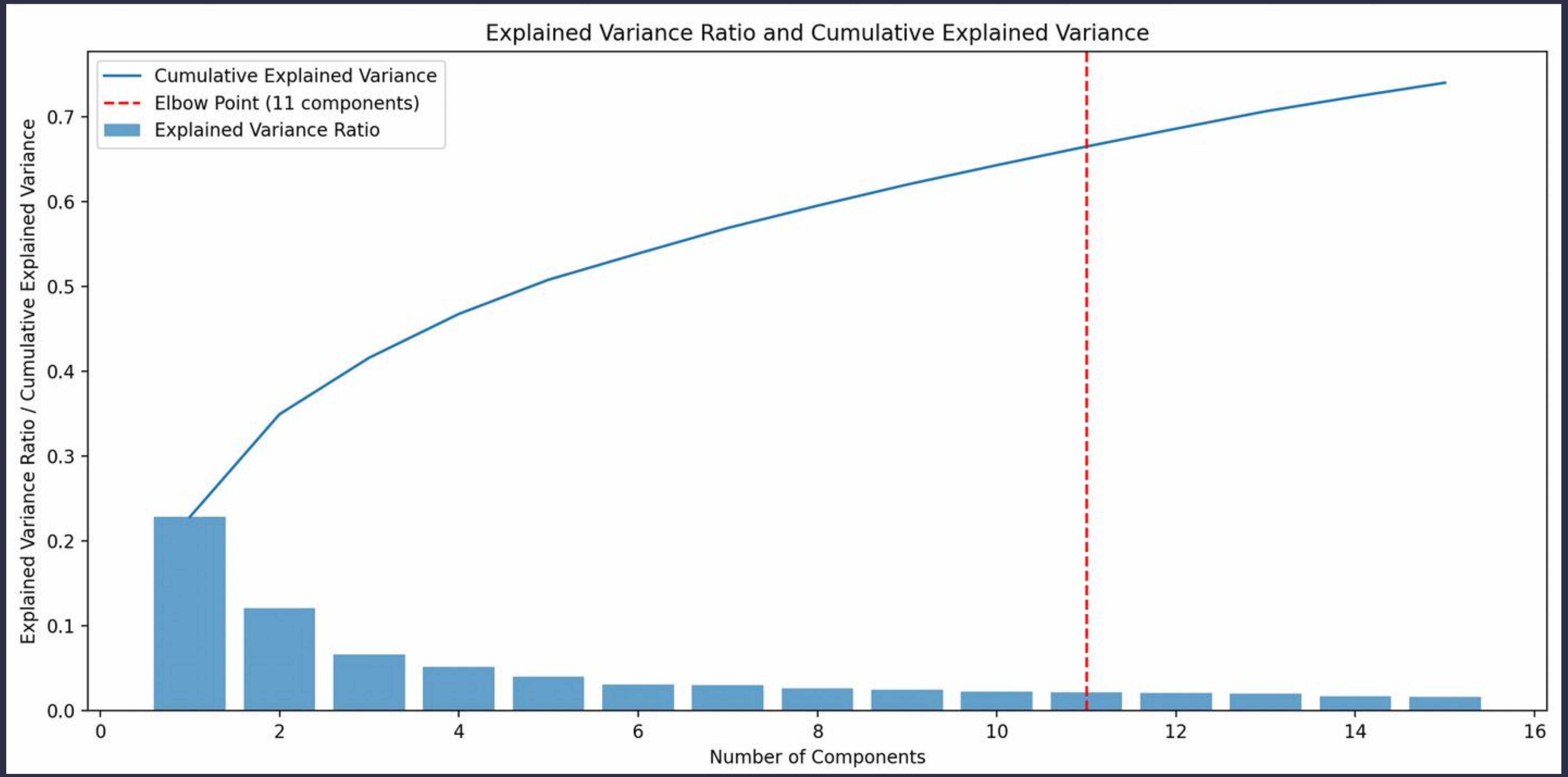


Features for De-noised waveforms

Audio

Step 4: Dimensionality Reduction

- Using PCA
- Normalize the values
- Save the new df along with truth labels.

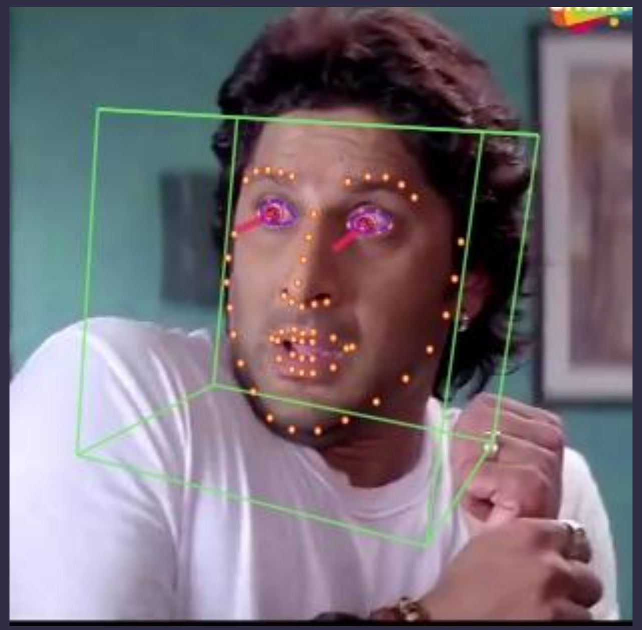


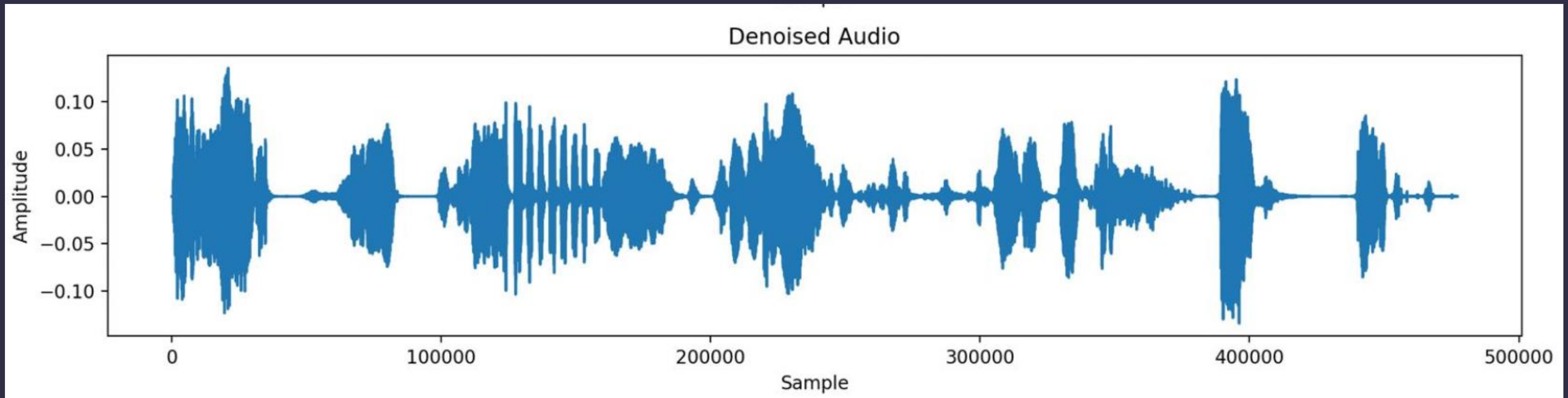
PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	usernum	run	truth
0.3565884195868340	0.3142881269533760	0.2089309688751570	0.354409220949523	0.1865248157888450	0.2811003257612250	0.3850417481625030	0.5560918129382020	0.6780761667874550	0.7582474905883730	0.4436708500343090	0	1	0
0.6653900899091070	0.5065298031504780	0.5296070284116260	0.4152051955349910	0.1516482384020340	0.2459110762712520	0.3059950101818420	0.5958016329401690	0.8473613976477100	0.6428338272716260	0.3694720077501210	0	2	0
0.5662026494266750	0.3806785894860090	0.2182968064603390	0.4614453178475810	0.3036362642432760	0.4829758137133170	0.4320814570603540	0.4213877103496990	0.7892140879720750	0.513877143942343	0.4452422915090130	0	3	0
0.5471616466276580	0.3300928940796350	0.2835637729689690	0.6387943215095100	0.3091860316414560	1.0	0.7592998793618940	0.4846360734489530	0.3063976098830150	0.5532122062143120	0.6474007661810810	0	4	0
0.5198656431696860	0.3114621701590690	0.1385635023235170	0.4745563258175360	0.3937405073200040	0.4453349638621760	0.4360784940230950	0.353628671886611	0.5646628295484870	0.3682325944095700	0.3827941177578940	0	5	1
0.5846281992486940	0.2813520772052210	0.1366728224732310	0.2636855337257330	0.2811559398582290	0.3675508148197260	0.5646095784504610	0.5362404035854540	0.7758118240482830	0.3061666536527180	0.4141574304346040	1	0	1
0.5194256263979070	0.3426243223964160	0.2388022679424180	0.4283839012331230	0.3656075607536190	0.5393257618460670	0.4077606766686450	0.5284278253306370	0.8677410323431770	0.2567730966191660	0.5284468663032770	1	1	0
0.4715557970231300	0.3063179705256300	0.1276246725851110	0.3048736779528260	0.3271926224568070	0.4507266939779670	0.4764016034551090	0.6340713156780560	0.7964518039948490	0.1883748696998950	0.5385219168604280	1	4	0
0.3584691322780460	0.3901258499218200	0.1337462986455400	0.1737834320000050	0.4178770242347590	0.3691307761566720	0.4239442073646480	0.4357521235938100	0.556319306196135	0.3032522593613860	0.562520815039449	1	5	1
0.4655228955172690	0.5275532780372670	0.3766745594882340	0.6372396936632510	0.267971338045578	0.4439666864265730	0.5813645601883060	0.2282034072587570	0.5610516788655860	0.5506544608298070	0.3692330754967290	10	2	1
0.4131371821193270	0.4653349303225280	0.4713279545469760	0.6274569897877240	0.3172781945642900	0.5436246920975980	0.4491226550947710	0.3475385401024690	0.3882792390108620	0.6087982234481890	0.4405018014725140	10	3	1
0.3503095583288230	0.64274928421437	0.4556876363566360	0.7082816683958980	0.3250669250115550	0.6289188077604010	0.4506141811867550	0.3186477317860680	0.5814206462178850	0.6911875627238830	0.4317167960785540	10	4	0

shape: 14 x 205

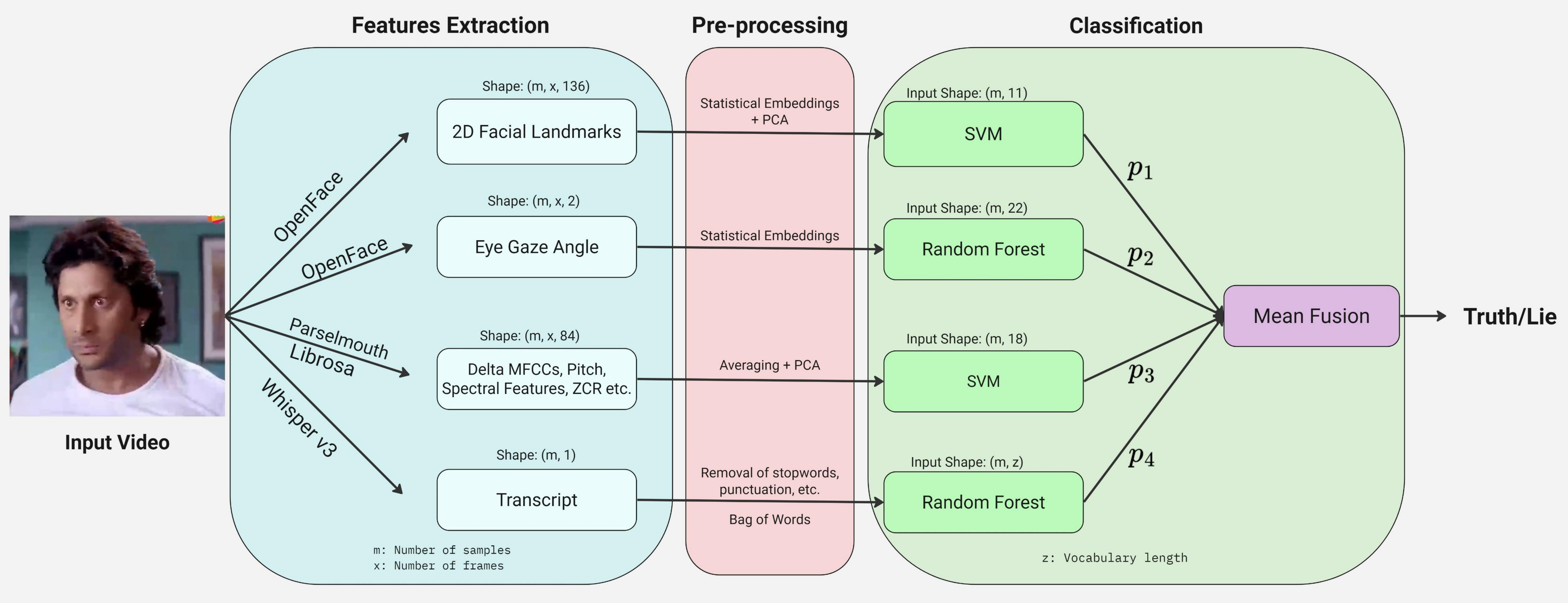
Input Modalities

Transcript: Aunty chooha!





ML Methodology



Our model

Support Vector Classifier

- Support vectors - similar expression/audio features
- Based on statistical embeddings
- Works well for small sample size

Random Forest Classifier

- Avoids overfitting
- Gaze and text have limited domain/vocabulary

Late Fusion technique

- Averages the output probabilities from every modality
- Different human behaviours are captured across each modality.

Challenges faced:

Challenge: Bag of Lies (primary dataset) was made accessible in March
Solution: Worked with Court Trial dataset for initial feature extraction

Challenge: Limited number of samples in dataset for applying NN/LSTM
Solution: Used other classifiers with similar baseline performance

Challenge: CUDA compatibility issues + server crashes
Solution: Re-installed our operating systems

Results

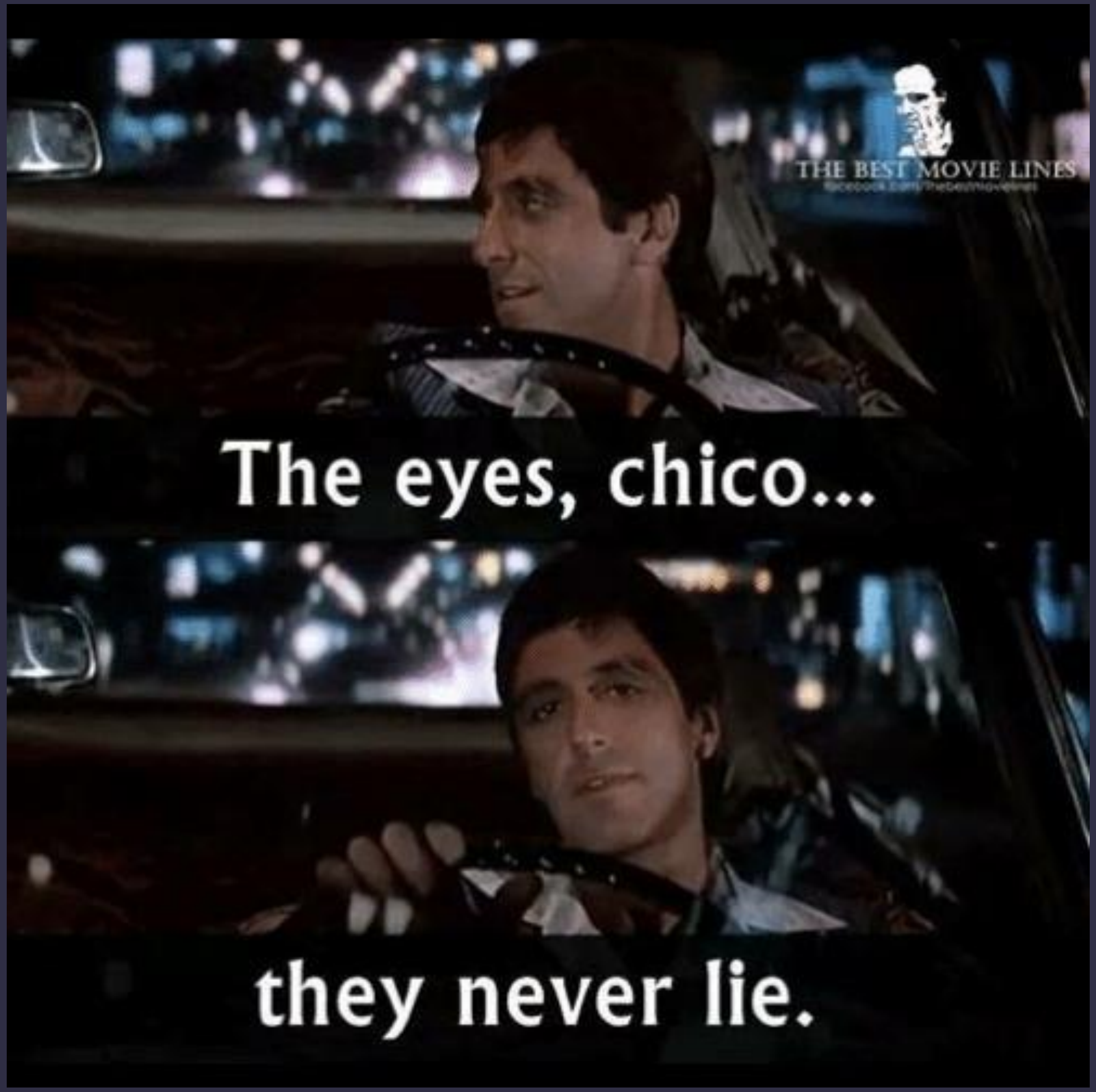
		Set A - Audio + Video (all 325)						Set B - Audio + Video + Transcripts (205)					
Modalities	**Model**	**1**	**2**	**3**	**4**	**5**	**Mean Accuracy**	**1**	**2**	**3**	**4**	**5**	**Mean Balanced Accuracy**
Only Audio	SVM (RBF)	0.6123542	0.6123542	0.6123542	0.6123542	0.6123542	**0.6123542**	0.5630272	0.5630272	0.5630272	0.5630272	0.5630272	0.563027211
	Random Forest	0.5823181	0.5823181	0.5823181	0.5823181	0.5823181	0.5823181	0.5496939	0.5496939	0.5496939	0.5496939	0.5496939	0.5496939
	Gaussian NB	0.5715461	0.5715461	0.5715461	0.5715461	0.5715461	0.5715461	0.5728571	0.5728571	0.5728571	0.5728571	0.5728571	**0.5728571**
Only Video (2D Facial Landmarks + PCA)	Neural Network	0.5457524	0.5307761	0.5347784	0.5475778	0.5358122	0.5389394	0.556483	0.5475993	0.5654397	0.5543457	0.5467433	0.5541222
	SVM (Sigmoid)	0.5801175	0.5801175	0.5801175	0.5801175	0.5801175	**0.5801175**	0.567483	0.567483	0.567483	0.567483	0.567483	**0.567483**
	Random Forest	0.5381334	0.5147207	0.5135302	0.5144826	0.5416254	0.5244985	0.5412925	0.4894898	0.5035374	0.5336395	0.5098299	0.5155578
Only Transcripts	Gaussian NB	NA						0.6692517	0.6692517	0.6692517	0.6692517	0.6692517	0.6692517
	SVM (RBF)							0.6577211	0.6577211	0.6577211	0.6577211	0.6577211	0.6577211
	Random Forest							0.7042177	0.6731293	0.6907823	0.7105102	0.7067347	**0.6970748**
Only Gaze (averaged angle)	RF	0.6590641	0.6604927	0.6695187	0.6446774	0.6733282	**0.6614162**	0.5831633	0.5815986	0.5958844	0.5937415	0.5857823	**0.588034**
	SVM (poly)	0.5641084	0.5641084	0.5641084	0.5641084	0.5641084	0.5641084	0.555102	0.555102	0.555102	0.555102	0.555102	0.555102
	KNN (n = 15)	0.6009946	0.6009946	0.6009946	0.6009946	0.6009946	0.6009946	0.5454422	0.5454422	0.5454422	0.5454422	0.5454422	0.5454422
Video + Audio	Early Fusion	0.5570367	0.5570367	0.5570367	0.5570367	0.5570367	0.5570367	0.5471429	0.5471429	0.5471429	0.5471429	0.5471429	0.5471429
	Late Fusion	0.5931117	0.5842228	0.5845403	0.5776139	0.5921593	**0.5863296**	0.5535714	0.5535714	0.5535714	0.5535714	0.5535714	**0.5535714**
Video + Transcripts	Early Fusion	NA						0.5471429	0.5471429	0.5471429	0.5471429	0.5471429	0.5471429
	Late Fusion							0.580102	0.5683673	0.5748299	0.587585	0.5752041	**0.5772177**
Video + Gaze	Early Fusion	0.5598938	0.5598938	0.5598938	0.5598938	0.5598938	0.5598938	0.5471429	0.5471429	0.5471429	0.5471429	0.5471429	0.5471429
	Late Fusion	0.6564667	0.6564667	0.6564667	0.6564667	0.6564667	**0.6564667**	0.5788776	0.5815306	0.5629592	0.6223469	0.6012925	**0.5894014**
Audio + Transcripts	Early Fusion	NA						0.5541497	0.5541497	0.5541497	0.5541497	0.5541497	0.5541497
	Late Fusion							0.5466667	0.5663265	0.5606122	0.5663265	0.5642857	0.5608435
Audio + Gaze	Early Fusion	0.6579674	0.6579674	0.6579674	0.6579674	0.6579674	**0.6579674**	0.6097959	0.6097959	0.6097959	0.6097959	0.6097959	**0.6097959**
	Late Fusion	0.6455143	0.6383498	0.6697784	0.6775922	0.6432849	0.6549039	0.5977551	0.5822789	0.6010884	0.5902721	0.6147959	0.5972381
Gaze + Transcripts	Early Fusion	NA						0.5453741	0.5453741	0.5453741	0.5453741	0.5453741	0.5453741
	Late Fusion							0.6445578	0.6068367	0.6136054	0.6240816	0850340136	**0.6222704**
Audio + Video + Transcripts	Early Fusion	NA						0.5471429	0.5471429	0.5471429	0.5471429	0.5471429	0.5471429
	Late Fusion							0.5680272	0.5591837	0.5620748	0.5659864	0.5605442	**0.5631633**
Video + Gaze + Audio	Early Fusion	0.5598938	0.5598938	0.5598938	0.5598938	0.5598938	0.5598938	0.5471429	0.5471429	0.5471429	0.5471429	0.5471429	0.5471429
	Late Fusion	0.6748145	0.6592806	0.651878	0.6549948	0.6423758	**0.6566687**	0.6018707	0.6156122	0.6205442	0.6132313	0.5923129	**0.6087143**
Video + Gaze + Transcripts	Early Fusion	NA						0.5471429	0.5471429	0.5471429	0.5471429	0.5471429	0.5471429
	Late Fusion							0.604932	0.6105442	0.6126531	0.6153061	0.6060204	**0.6098912**
Audio + Gaze + Transcripts	Early Fusion	NA						0.5471429	0.5471429	0.5471429	0.5471429	0.5471429	0.5471429
	Late Fusion							0.604932	0.6132653	0.6117347	0.6016667	0.6163265	**0.609585**
Audio + Video + Transcripts + Gaze	Early Fusion	NA						0.5471429	0.5471429	0.5471429	0.5471429	0.5471429	0.5471429
	Late Fusion							0.5901361	0.5993197	0.579932	0.6061224	0.6061224	**0.5963265**

Different models tried with different modalities

Results

Modalities	Model	Set A (325 samples)	Set B (205 samples)
Only Audio	SVM (rbf)	**61.24%**	56.30%
	Random Forest	58.23%	54.97%
	GaussianNB	57.15%	**57.27%**
Only Video	Neural Network	53.89%	55.41%
	SVM (sigmoid)	**58.01%**	**56.78%**
	Random Forest	52.45%	51.56%
Only Text	Gaussian NB	-	66.93%
	SVM (RBF)		65.77%
	Random Forest		**69.53%**
Only Gaze	Random Forest	**66.14%**	**58.80%**
	SVM (poly)	56.41%	55.51%
	KNN (n = 15)	60.10%	54.54%
Video + Audio	Mean Fusion	58.64%	55.36%
Video + Transcripts	Mean Fusion	-	57.72%
Video + Gaze	Mean Fusion	65.64%	58.94%
Audio + Transcripts	Mean Fusion	-	56.08%
Audio + Gaze	Mean Fusion	65.80%	59.72%
Gaze + Transcripts	Mean Fusion	-	62.22%
Audio + Video + Transcripts	Mean Fusion	-	56.31%
Video + Gaze + Audio	Mean Fusion	65.67%	60.87%
Video + Gaze + Transcripts	Mean Fusion	-	60.99%
Audio + Gaze + Transcripts	Mean Fusion	-	60.968
All four	Mean Fusion	-	59.63%

Best performing models with late fusion



Scarface
(1983)

Results

Set B - LOGOCV Balanced Accuracy

- 40-60 class imbalance
- Highest accuracy: 69.53%
- No literature data available on Set B

Set A - LOGOCV Accuracy

- Highest accuracy: 66.14% (Gaze)
- On par with other research
- Does not use expensive equipments

Modality	Method	Average Accuracy	
		Set A	Set B
Only EEG	Random Forest	58.71	-
	EEG Net	54.25	-
	MLP	53.79	-
Only Gaze	Random Forest	61.70	57.11
	MLP	57.71	53.51
Only Video	LBP + SVM	55.21	53.25
	LBP + Random Forest	56.20	55.26
	LBP + MLP	54.22	49.90
Only Audio	Random Forest	53.24	54.89
	KNN	53.22	56.22
EEG + Gaze	Score level fusion of best performing algorithms on various modalities	62.22	-
EEG + Audio		61.69	-
EEG + Video		60.20	-
Gaze + Audio		63.69	59.42
Gaze + Video		62.19	62.71
Audio + Video		60.68	58.24
Gaze + Video + EEG		62.70	-
Gaze + Audio + EEG		63.21	-
Audio + Video + EEG		63.18	-
Gaze + Video + Audio		64.69	60.09
All four		66.17	-

Bag-of-Lies 11-fold accuracy without EEG (Set A): **64.69%**

In Plaksha

To check for Academic disintegrity

We hope that you will uphold the highest standards of academic integrity and refrain from any behaviour that compromises these standards.

Best Wishes!

The Office of Academic Affairs

Can be used by the Disciplinary committee

To detect Substance consumption by students

Thank You :)