Sound Spaces

Xavier Nishikawa, Ellie Kung, Matthew Kwon
Quantitative Engineering Analysis 1, Fall 2024

Abstract

The goal of this project is to develop a speaker recognition system that uses Fisherfaces to classify and group voice spectrograms based on similarities in how different individuals pronounce the same name.

I. Process

1. Collect audio data of people saying the same name.
2. Convert audio recordings into a spectogram.
3. Use fisher faces and linear discriminant analysis to reduce dimensions.
4. Use this found data as a training set and get another person to say the names as a test set.
5. See if we are able to properly differentiate each name when it is said by different people

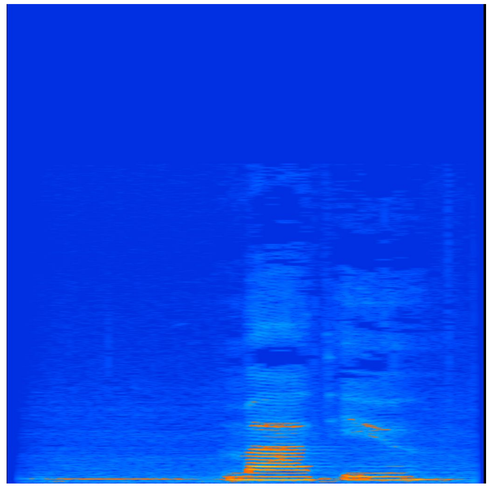


Fig 1. A spectrogram of a subject saying “Matthew”

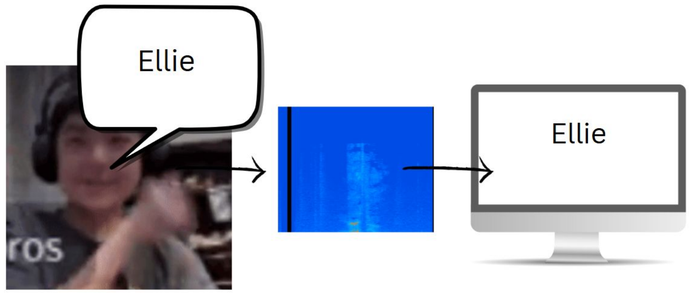
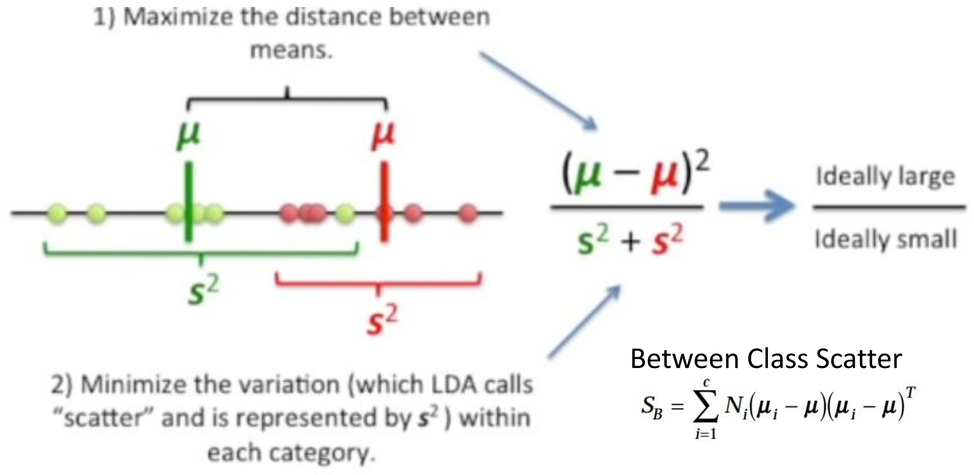


Fig 2. The process of our paper.

II. Explaining LDA



Between Class Scatter

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T$$

Within Class Scatter

$$S_W = \sum_{i=1}^{c} \sum_{\mathbf{x}_k \in X_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^T$$

III. How Fisher Faces Work

Fisherfaces combine PCA and LDA to create features that maximize class separation. PCA reduces dimensionality, and LDA optimizes the data by minimizing within-class variance and maximizing between-class variance, making Fisherfaces highly effective for classification tasks.

$$W_{opt}^T = W_{fld}^T W_{pca}^T$$

$$W_{pca} = \arg\max_W \left| W^T S_T W \right|$$

$$W_{fld} = \arg\max_W \frac{\left| W^T W_{pca}^T S_B W_{pca} W \right|}{\left| W^T W_{pca}^T S_W W_{pca} W \right|}$$

Fig 3. The equations Fisher came up with to optimize projection of Fisher faces.

IV. Minimizing Class Variability

Fig 4. Example of the different lighting

Fisherfaces can account for the variance in lighting as it minimizes the class variability, making an average face for the class. In our case, it would minimize the variability in voices, like pitch, tone, and volume

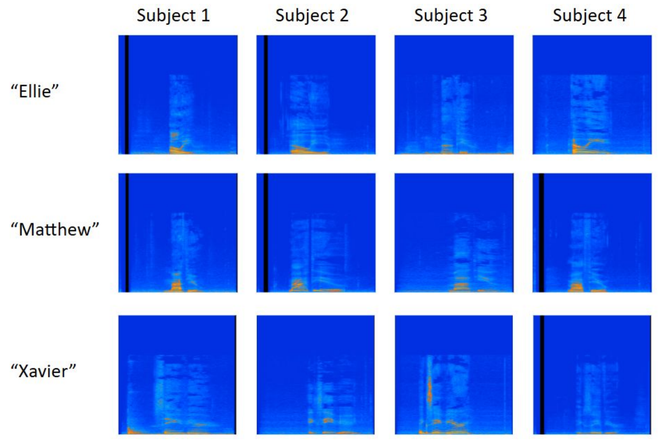


Fig 5. Some of our dataset which shows variance within each subject’s voice

V. Next Steps

We would like to collect more voices to account for a wider variation in microphone distance and speaking tones and automate the process of turning the recordings into spectrograms.

VI. References

Belhumeur, P.N., et al. “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, July 1997, pp. 711–720, https://doi.org/10.1109/34.598228.

StatQuest with Josh Starmer. “StatQuest: Linear Discriminant Analysis (LDA) Clearly Explained.” *Www.youtube.com*, 10 July 2016, www.youtube.com/watch?v=azXCzI57Yfc&ab_channel=StatQuestwithJoshStarmer.