EECS 16A
Voice Recognition 2

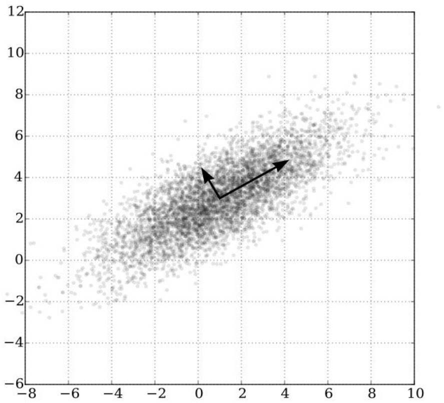
Welcome! We'll be starting at Berkeley Time.

Today's Agenda

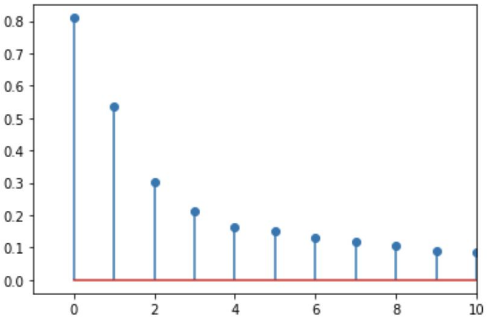
- SVD / PCA Review
- Revisiting Spectral Analysis
  - STFT & Spectrograms
- Reducing STFT results into vectors
- Experimentation

PCA Review

- PCA = Principal Component Analysis
  - **Principal components:** basis vectors that maximize variance in our data
  - Oftentimes, we can capture most of the data's behavior with just a few principal components!
    - **Fewer dimensions is easier to work with**
- How do we compute PCA?
  - Let's use SVD!!!!
  - Take the vectors that correspond with the highest singular values since those are the "most important" transformations of a matrix
  - The principal components of our setup are the vectors from V (basis for rowspace of A)



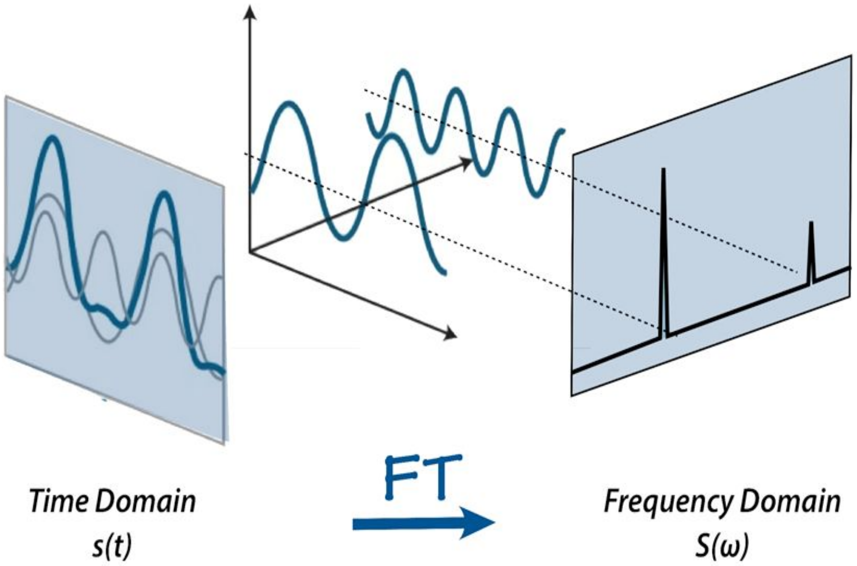
Principal Components of Data Example



Sigma Values Example

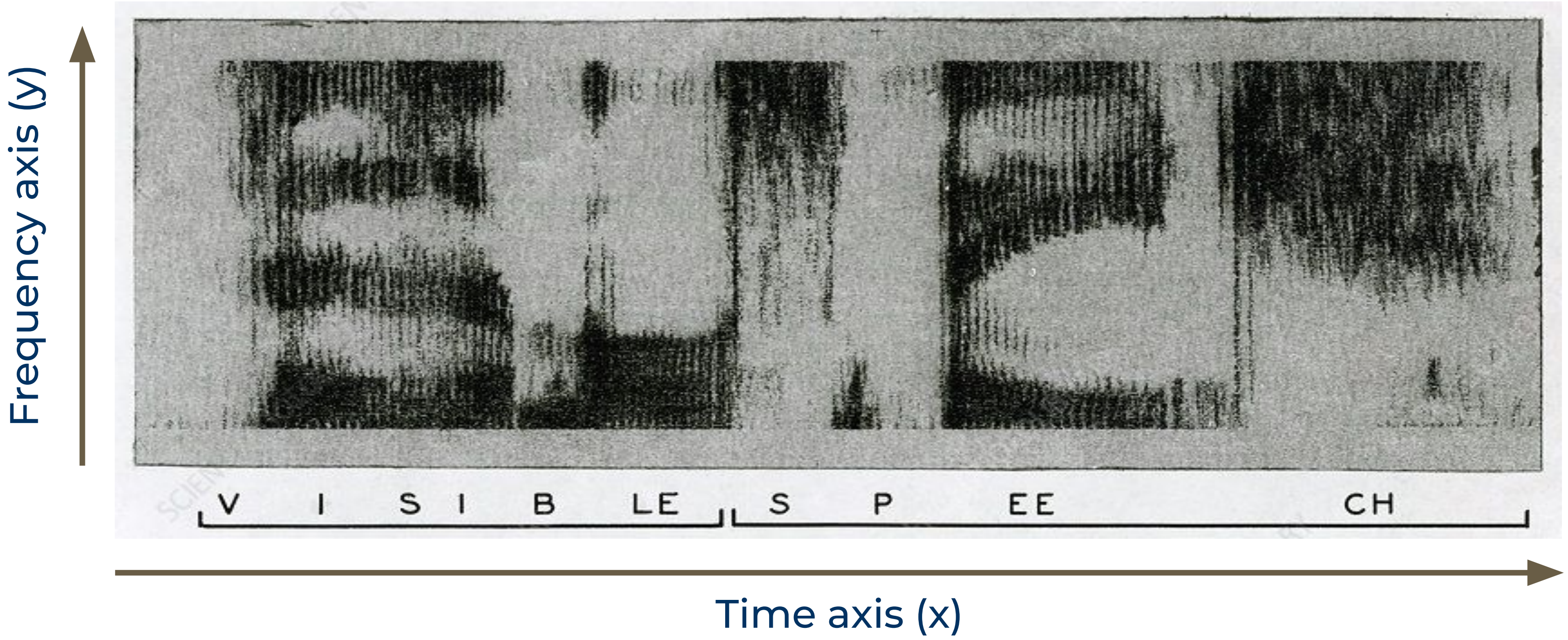
DFT Review

- DFT is a way to turn a time-domain signal into a frequency-domain spectrum
- The spectrum tells you about the frequency content of the signal



Spectrograms and STFT

- If you do a DFT on many short slices of the signal, you get a spectrogram (temporal information)



Spectrograms and STFT

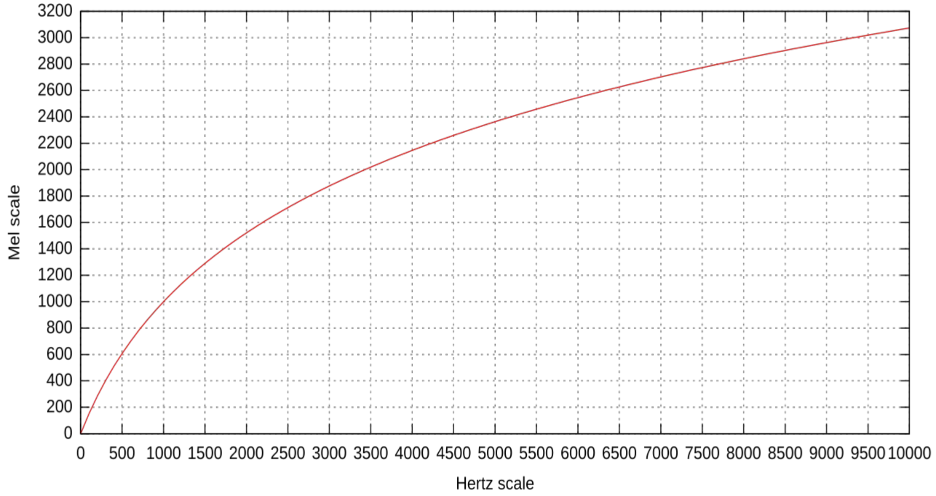
- The STFT (Short-time Fourier Transform) is essentially generating a spectrogram, but a spectrogram by definition is the magnitude-squared STFT:

$$\mathrm{spectrogram}(t, w) = |\mathrm{STFT}(t, w)|^2$$

- The STFT may result in a complex-valued spectrum, but the spectrogram is always real-valued. We will be taking the absolute value of the STFT in this lab.

Mel-scaled STFT

- Mel scale: Frequency scaling based on human pitch perception
- We can calculate the STFT where the frequency bins are apart by units in Mels instead of Hz



Voice Classification with PCA

Our voice classification system from last week:

1. Data Pre-Processing
2. SVD and PCA Computation
3. Mean Centroid Classification
4. Validation + Hyperparameter Tuning

Voice Classification with STFT+PCA

Our voice classification system for this week:

1. Data Pre-Processing
2. **Compute STFT of audio signal**
3. **Reduce STFT (2D matrix) into 1D vector**
4. SVD and PCA Computation
5. Mean Centroid Classification
6. Validation + Hyperparameter Tuning

Can we use the STFT result directly?

- Our STFT result is 2D (a matrix)!
- Q: Can we use this result in our PCA scheme? What input do we expect?
- A: We expect a vector for our PCA scheme. Thus, we need to squish our STFT result down to a vector somehow while retaining information.

Reduction Method 1: Flattening

- We can simply flatten our STFT result matrix into one long vector. This results in a really long vector but we don't lose any information!

$$\begin{bmatrix} S_{11} & S_{12} & S_{13} & \dots \\ S_{21} & S_{22} & S_{23} & \dots \\ \vdots & & \ddots & \end{bmatrix} \Longrightarrow$$

$$\begin{bmatrix} S_{11} & S_{12} & S_{13} & \dots & S_{1N} & S_{21} & S_{22} & S_{23} & \dots \end{bmatrix}$$

Reduction Method 2:
Aggregation Using SD and Variance

- We can turn the vector into a more compact form by only saving the standard deviation and variance of each time slice in the STFT result as well.

$$\begin{bmatrix} S_{11} & S_{12} & S_{13} & \cdots \\ S_{21} & S_{22} & S_{23} & \cdots \\ \vdots & & \ddots & \end{bmatrix} \Longrightarrow \begin{bmatrix} \mu_{S1} \\ \sigma_{S1} \\ \mu_{S2} \\ \sigma_{S2} \\ \vdots \end{bmatrix}$$

Experimentation

- We have 4 different possibilities we could try:
  - STFT + Flattening
  - Mel-scaled STFT + Flattening
  - STFT + SD/Variance Aggregation
  - Mel-scaled STFT + SD/Variance Aggregation
- You are encouraged to experiment between these four configurations!
- In order to try different methods, you will have to change:
  1. processed_A used in the training step
  2. processed_A_test used in the test step
  3. STFT / reduction functions used in the live classification
- **Try at least two different methods, and reason about the differences during checkoff.**

Feedback

Please provide feedback with this anonymous feedback form!!

https://tinyurl.com/fb-student-fa24