

DSP LAB

REAL-TIME TEXT-INDEPENDENT SPEAKER IDENTIFICATION

Presented by Shihong Fang and He Huang

[illegible]

A stylized illustration of a human head in profile, facing left. The head is white with a thick black outline. Inside the head, there are several blue gears of different sizes. The largest gear is in the center, with the word 'idea' written in a small, black, sans-serif font inside its central circle. Other gears are scattered around it, some overlapping. The background is a solid light blue. In the top left corner, there are faint, light blue icons: a lightbulb, a magnifying glass, and a gear. In the top right corner, there is a small, light blue star and a faint outline of a document or book.

LAB

AL-TIME

DEPENDENT

IDENTIFICATION

ng and He Huang



Speaker Identification

In speaker identification, the goal is to determine which one of a group of known voices best matches the input voice sample. Furthermore, in either task the speech can be constrained to be a known phrase (text-dependent) or totally unconstrained (text-independent). Success in both tasks depends on extracting and modeling the speaker-dependent characteristics of the speech signal which can effectively distinguish one talker from another.

LAB

AL-TIME

DEPENDENT

IDENTIFICATION

ng and He Huang



Ideas

Training

1. Extract human voice features
2. Build models for each person based on the features extracted

Testing

1. Listen to the voice people's speaking and extract features
2. Apply models in the database and see which person's models best fit the features
3. Find out the speaker

Extract voice features(MFCC)

1. Frame the signal into short frames.
2. For each frame calculate the periodogram estimate of the power spectrum.
3. Apply the mel filterbank to the power spectra, sum the energy in each filter.
4. Take the logarithm of all filterbank energies.
5. Take the DCT of the log filterbank energies.
Keep DCT coefficients 2-13, discard the rest.

Modeling

Gaussian Mix Model

Advantages:

Gaussian mixture models can represent general speaker-dependent spectral shapes and model arbitrary densities.

$$p(\vec{x}|\lambda) = \sum_{i=1}^M p_i b_i(\vec{x})$$

where \vec{x} is a D- dimensional random vector, $b_i(\vec{x})$, $i = 1, \dots, M$, are the component densities of D-variate Gaussian function and p_i , $i = 1, \dots, M$, are the mixture weights.

Algorithm

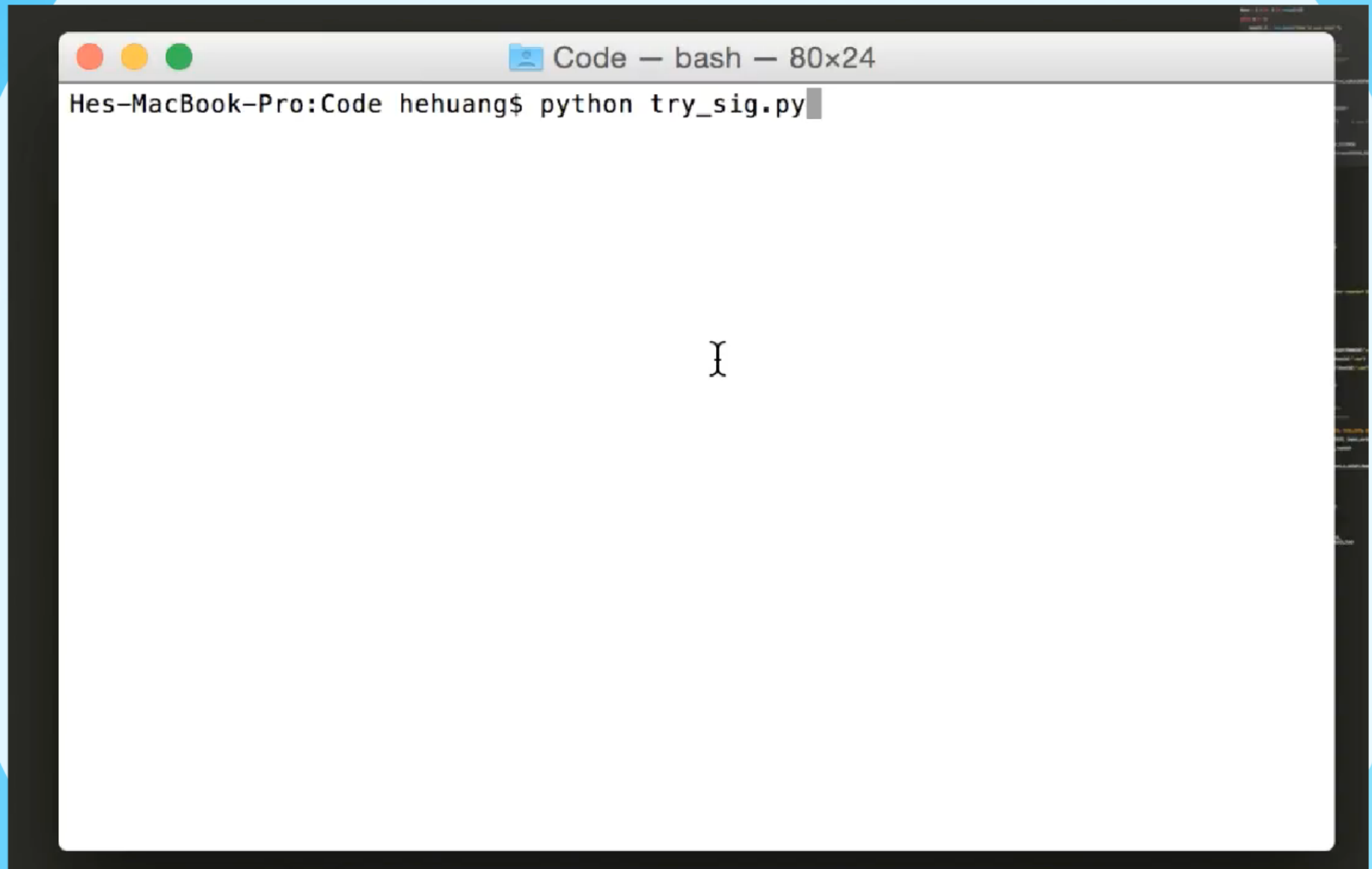
For a sequence of T training vectors $X = \{\vec{x}_1, \dots, \vec{x}_T\}$, the GMM likelihood can be written as

$$p(X|\lambda) = \prod_{t=1}^T p(\vec{x}_t|\lambda)$$

For speaker identification, a group of S speaker $\mathbf{S} = \{1, 2, \dots, S\}$ is represented by GMM's $\lambda_1, \lambda_2, \dots, \lambda_S$.

Using logarithms and the independence between observations, the speaker identification system computes

$$\hat{S} = \arg \max_{1 \leq k \leq S} \sum_{t=1}^T \log p(\vec{x}_t|\lambda)$$



```
Code — bash — 80x24
Hes-MacBook-Pro:Code hehuang$ python try_sig.py
```

Thank you !