DSP LAB

REAL-TIME TEXT-INDEPENDENT SPEAKER IDENTIFICATION

Presented by Shihong Fang and He Huang

DSP LAB

REAL-TIME TEXT-INDEPENDENT SPEAKER IDENTIFICATION

Presented by Shihong Fang and He Huang

P LAB

AL-TIME
DEPENDENT
DENTIFICATION

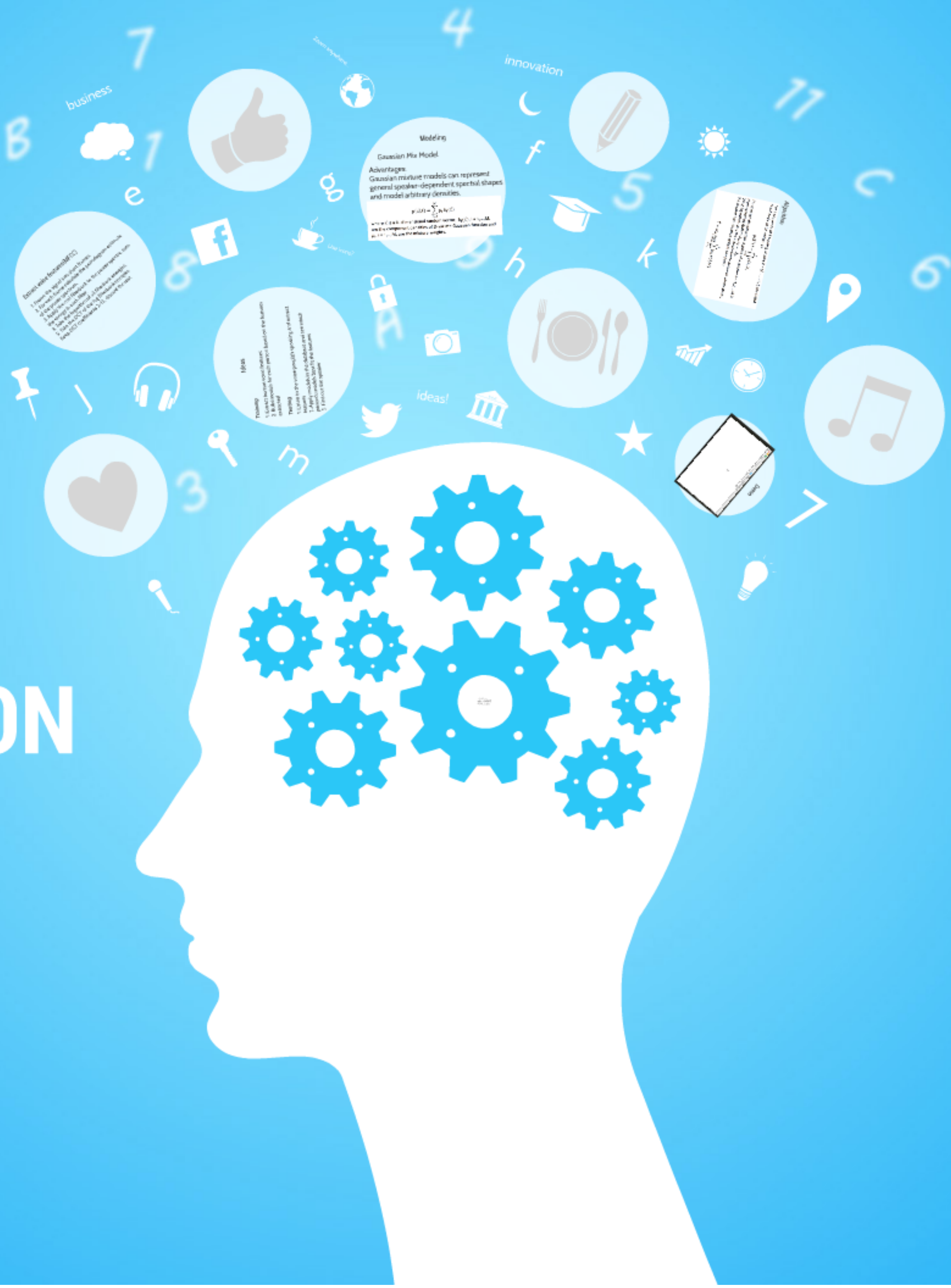
ng and He Huang

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.

P LAB

AL-TIME
DEPENDENT
DENTIFICATION



ng and He Huang

Prezi

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,…,M, are the component densities of D-variate Gaussian function and p_i, i = 1,…,M, are the mixture weights.

icons?

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 **S**= {1,2,...,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

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

Demo

Code — bash — 80×24

```
Hes-MacBook-Pro:Code hehuang$ python try_sig.py
```

Thank you !

Prezi