# MLLR on Emotional Speaker Recognition

Bie Fanhu

CSLT, RIIT, THU

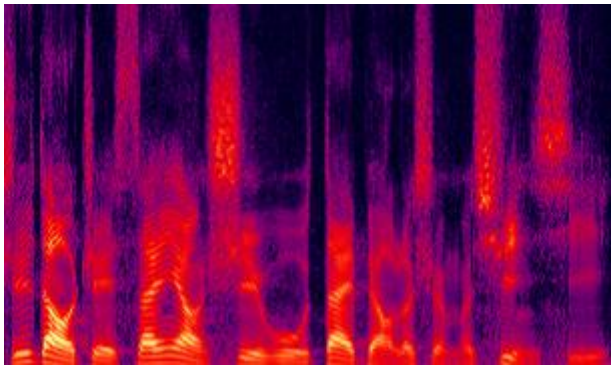
2013-05-06

# outline

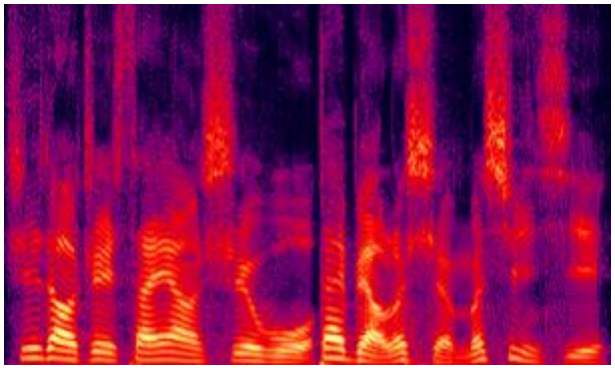
1. Spectrum Analysis for emotion utterances
2. Framework for GMM-UBM
3. Introduction for MLLR
4. MLLR for ESR
5. Experiments and results

# Emotional Utterances

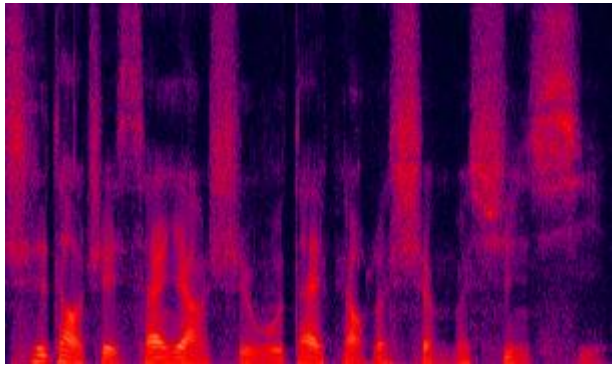
1. Spectrum for emotion utterances



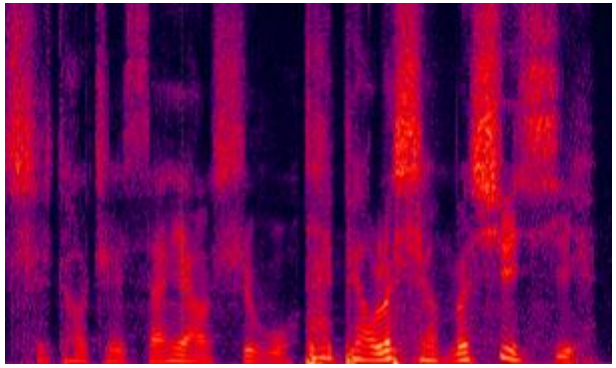
Neutral



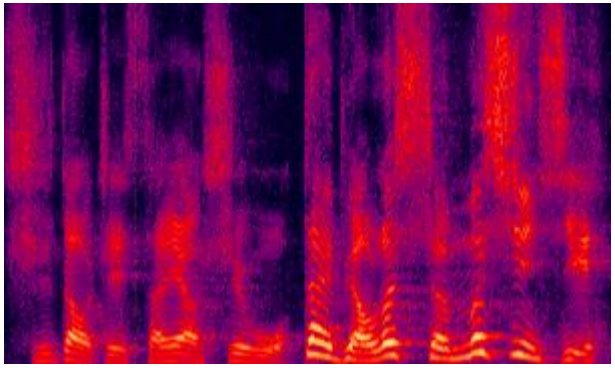
Happy



Sad



Anxious



Angry

# Emotional Utterances

1. Energy distribution

   - Concentrate
     - Happy high frequencies
     - Angry low frequencies
   - Distribute
     - Anxious high frequencies
     - Sad low frequencies

2. Time dimension

   The energy changes more in the end of the utterance

# GMM-UBM

1. Framework
   - UBM-Set of Gaussian mixtures
   - GMM
     - MAP from UBM
     - Gaussian mixtures maps to UBM one-to-one

2. Gaussian Mixture
   - Mean
   - Variance
   - Frequency
   - Energy

# MLLR

1. Introduction for MLLR

   - Element: Super vector
   - Type: Linear transformation
   - Target: Maximizing likelihood

2. MLLR mapping data

   - Source: models(GMM)
   - Destination: utterances or features

# MLLR

## 3. Simplified MLLR

- MLLRMEAN
  - Estimate the transformation for mean vector
  - Consider the covariance not changed

- Constrained MLLR
  - Estimate the transformation for both
  - Consider the mean and the variance share the same transformation

# MLLR for Emotional Speaker Recognition

1. MLLR is used to learn the difference of the energy distribution on different frequence
2. Each component is mapping to the same index of the component in UBM
3. MLLR can used to get the transformation for each component of the GMM
4. Each component share the same transformation

# MLLR for Emotional Speaker Recognition

## 5. Details

- Target: Transform the emotional space to the neutral space
- Source: Neutral models
- Destination: Emotional features
- Result: MLLR transformation

# Experiments and results

Condition 1: Some persons have enough emotional data, but the persons who will use the system don`t have emotional data for training.

Approach: The emotional data are used for training the MLLR matrix. The MLLR matrix is used to transform the features from neutral space to the emotional space

# Experiments and results

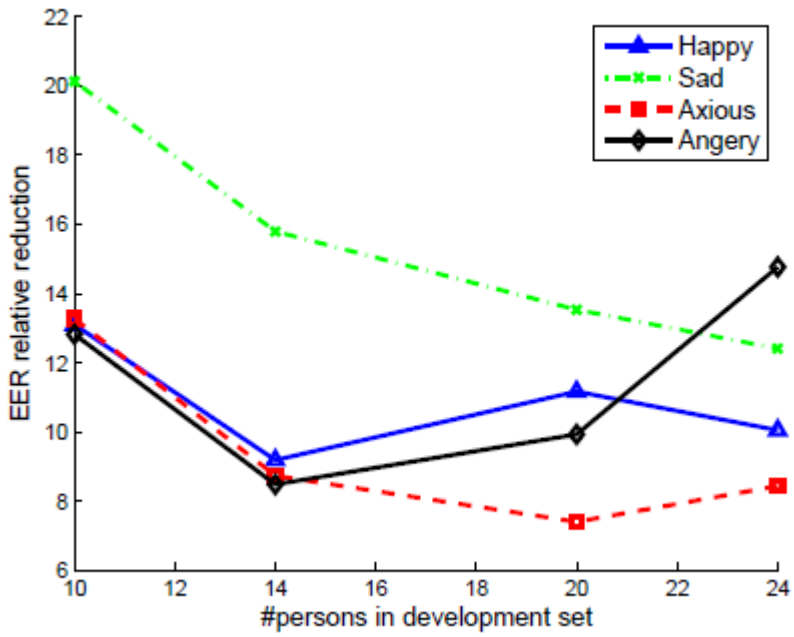
Data set trained for MLLR matrix contains 10 persons, totally 1000 utterances for each emtion, about 30 minutes

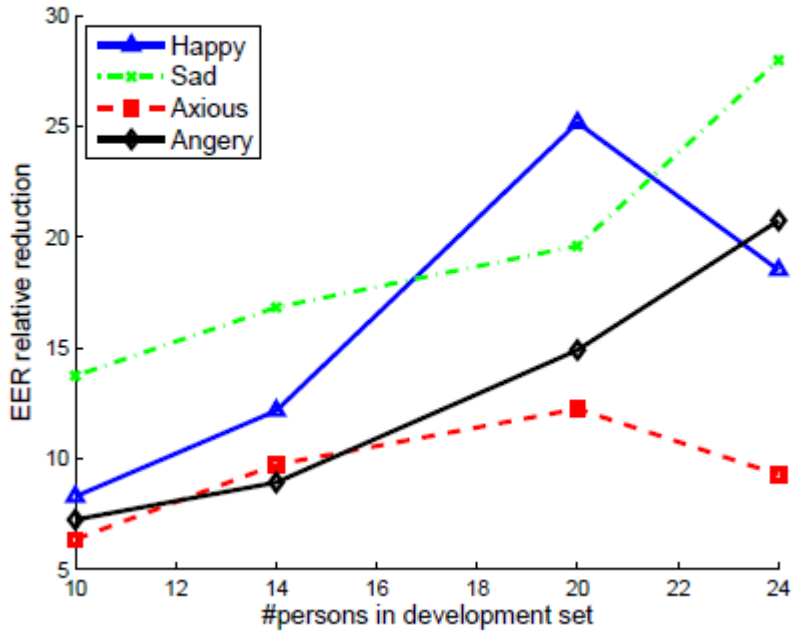
Results:

| | EER% | | |
|---|---|---|---|
| | Baseline | MLLR | CMLLR |
| Neutral | 2.15 | - | - |
| Happy | 14.61 | 13.36 | 12.71 |
| Sad | 20.16 | 17.05 | 16.55 |
| Anxious | 17.30 | 16.30 | 15.65 |
| Angry | 18.55 | 16.55 | 16.05 |

# Experiments and results

Different results when the train set increases

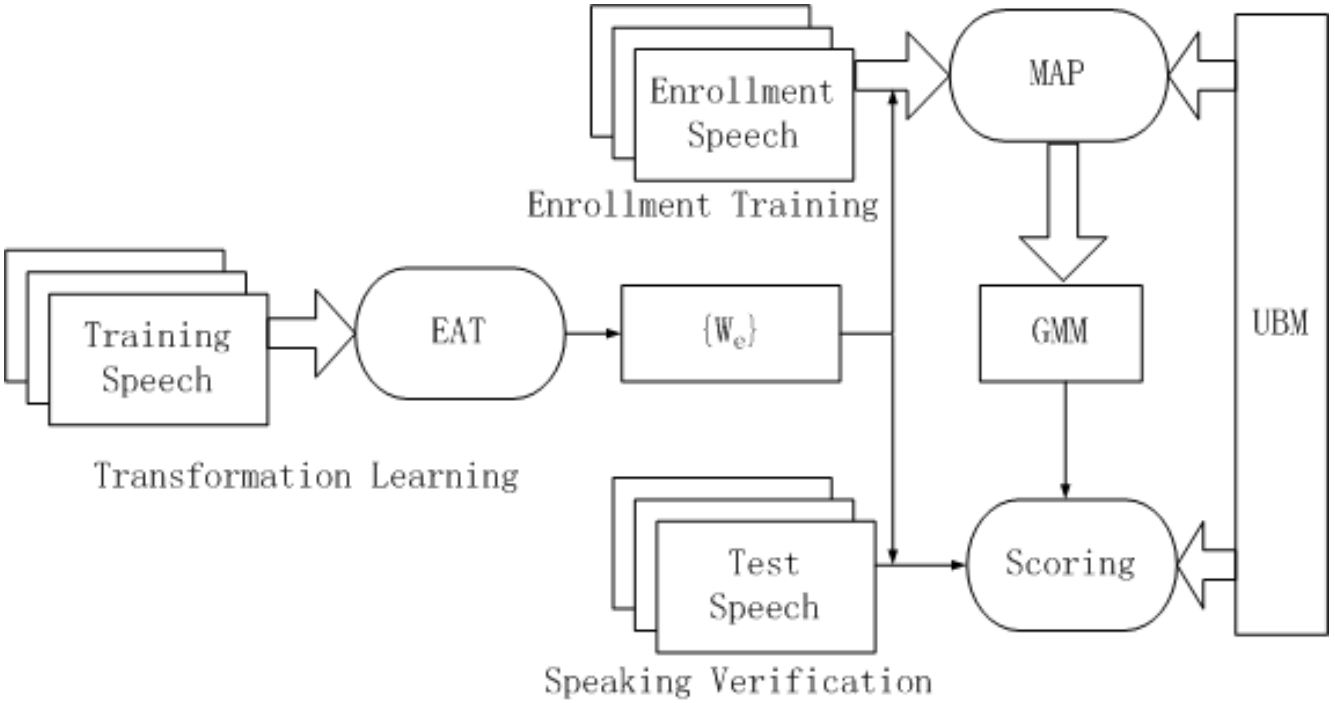


# Experiments and results

Condition 2: Some persons have enough emotional data, but the persons who will use the system don`t have enough emotional data for training.

# Approach: Emotional Adaptive Training



# Experiments and results

| | Train | Test |
|---|---|---|
| NMAP | Neutral data | Emotional data |
| EMAP | Emotional data | Emotional data |
| CMLLR | Neutral data | Emotional data |
| EAT-NE | Neutral data | Emotional data |
| EAT-EE | Emotional data | Emotional data |

# Experiments and results

Results:

| | EER% | | | | |
|---|---|---|---|---|---|
| | Neutral | Happy | Sad | Anxious | Angry |
| NMAP | 2.19% | 12.50% | 16.56% | 13.26% | 15.69% |
| EMAP | | 8.06% | 6.74% | 6.20% | 9.57% |
| CMLLR | | 10.50% | 14.94% | 12.39% | 14.20% |

| | EER% | | | |
|---|---|---|---|---|
| | Happy | Sad | Anxious | Angry |
| EAT-NE | 10.56% | 14.86% | 12.19% | 14.31% |
| EAT-EE | 5.38% | 6.36% | 5.37% | 7.88% |

# Experiments and results

Thank you!