Li Chen, Ying-chun Yang, Zhao-hui Wu, 2014. Mismatched feature detection with finer granularity for emotional speaker recognition. *Journal of Zhejiang University-SCIENCE C (Computers & Electronics)*, **15**(10):903-916.. [doi:10.1631/jzus.C1400002]

JZUS C (Comput & Electron)

Mismatched feature detection with finer granularity for emotional speaker recognition

Key words: Emotional speaker recognition, Mismatched feature detection, Feature regulation

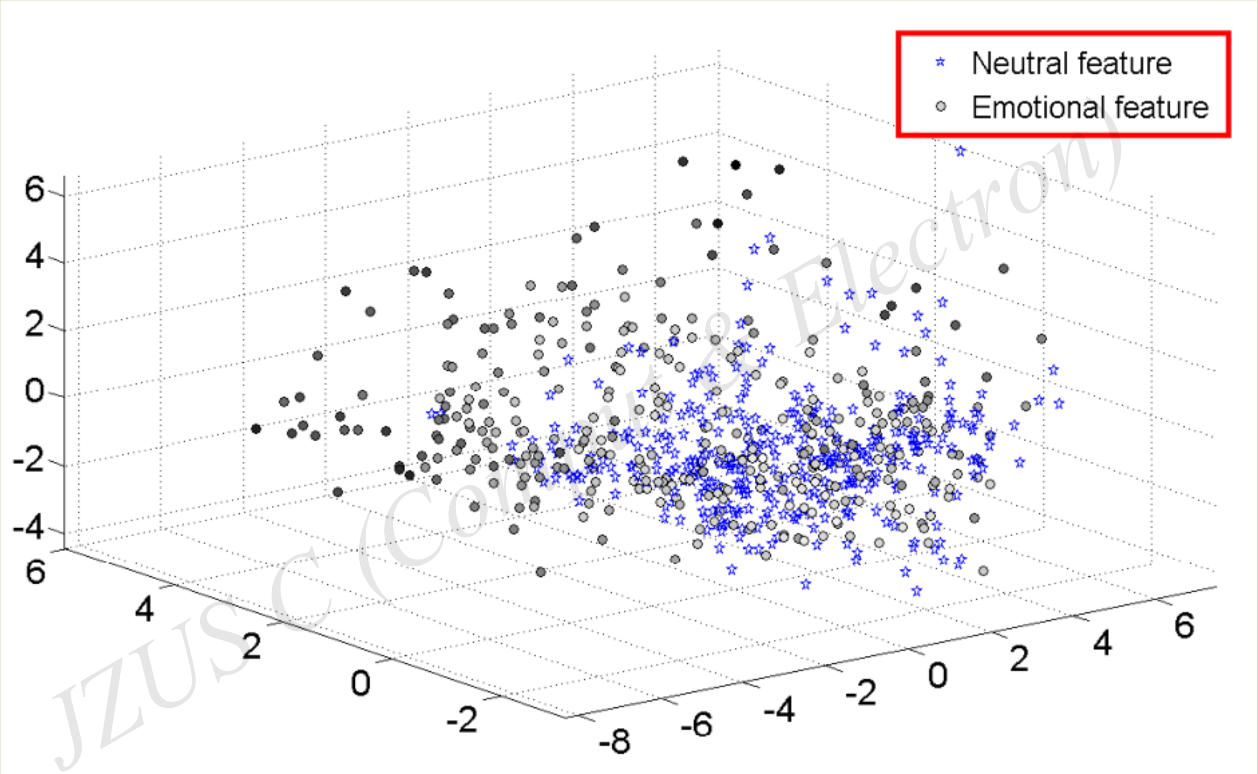
Corresponding author: Ying-chun Yang
E-mail: yyc@zju.edu.cn

Introduction

- The deviation of the emotional acoustic space of short-time speech features from the neutral acoustic space deteriorates the speaker recognition performance.
- Emotion variation produces different feature deformations for different phonemes.
- Feature pruning and feature regulation methods are proposed to process the mismatched features to improve the speaker recognition performance under each acoustic class.

IZUSC (Comput & Electron)

Mismatched feature



JZUS-C (Comput & Electron)

Mismatched features (represented by dark '•'), lying far away from the neutral acoustic space with weaker discriminative power, should be detected and regulated to mask their negative effects.

Mismatched feature detection

- Compute the occupation probability vector for each feature x_t under each component:

$$[r_1(x_t), r_2(x_t), \ldots, r_N(x_t)]$$

- Compute the matching score $RS_k(x_t)$ of the fuzzy SVM model of the *k*th component.

- Combine all matching scores $[RS_1(x_t), RS_2(x_t), \cdots, RS_N(x_t)]$ into the total matching score.

$$RS_t = \sum_{k=1}^{N} r_k(x_t) RS_k(x_t)$$

Those features with smaller matching scores are detected as mismatched features.

JZUS-C (Comput & Electron)

Mismatched feature regulation

- Aim:

$$\min \Phi_k, \quad \Phi_k = D_{w,k} - \alpha D_{b,k}$$

$$D_{w,k} = \sum_{s=1}^{S} \sum_{t=1}^{T_{s,e}} r_k(x_{s,t})(A_k x_{s,t} - \overline{x_s^k})^T (A_k x_{s,t} - \overline{x_s^k}), \quad x_{s,t} \in \mathbf{x_e}$$

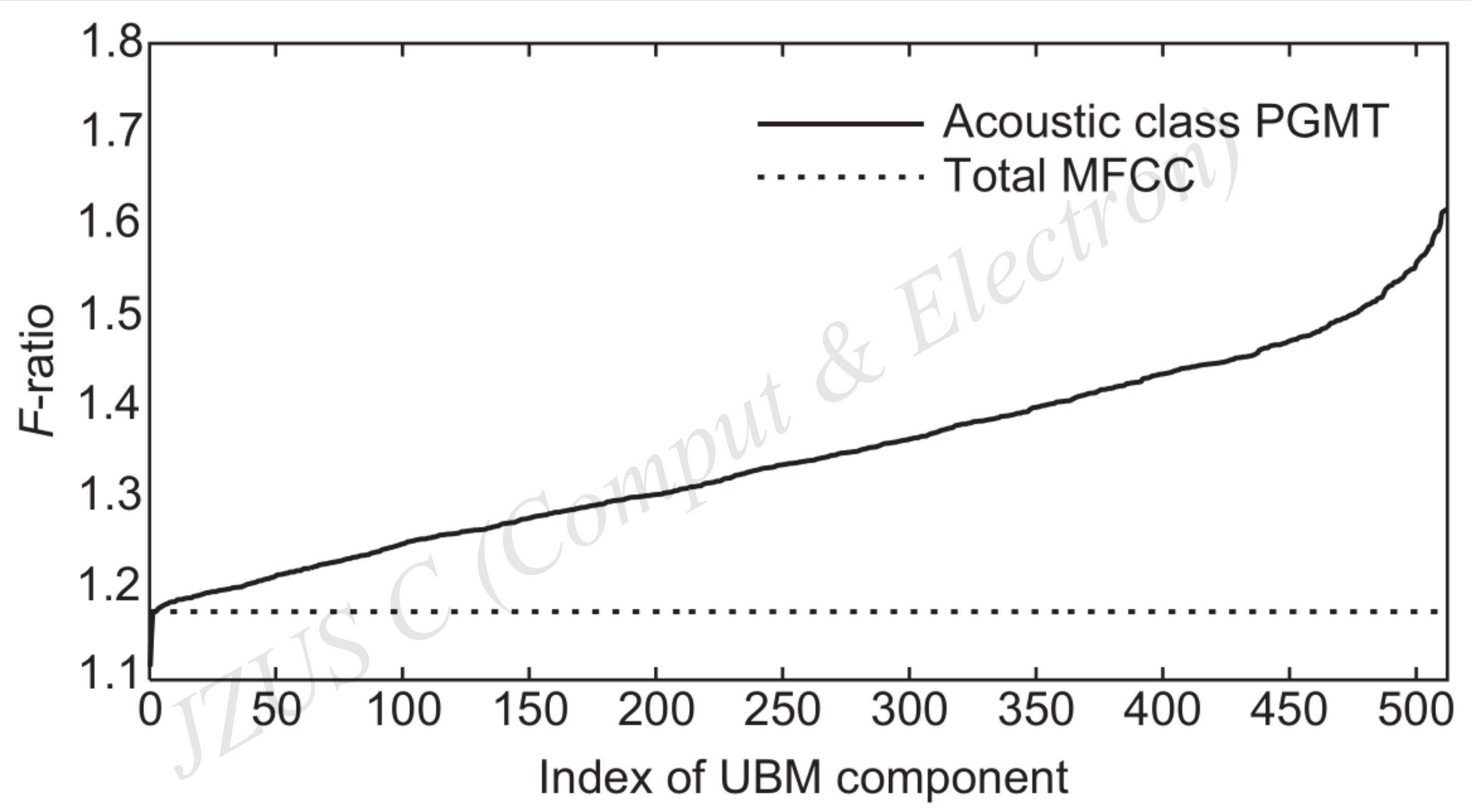
$$D_{b,k} = \frac{1}{S-1} \sum_{s=1}^{S} \sum_{t=1}^{T_{s,e}} \sum_{\hat{s}=1, \hat{s} \neq s}^{S} r_k(x_{s,t})(A_k x_{s,t} - \overline{x_{\hat{s}}^k})^T (A_k x_{s,t} - \overline{x_{\hat{s}}^k}), \quad x_{s,t} \in \mathbf{x_e}$$

- Solution:

$$A_k \sum_{s=1}^{S} \sum_{t=1}^{T_{s,e}} r_k(x_{s,t})(1-\alpha) x_{s,t} x_{s,t}^T = \sum_{s=1}^{S} \sum_{t=1}^{T_{s,e}} r_k(x_{s,t})(\overline{x_s^k} x_{s,t}^T - \frac{\alpha}{S-1} \sum_{\hat{s}=1, \hat{s} \neq s}^{S} \overline{x_{\hat{s}}^k} x_{s,t}^T).$$

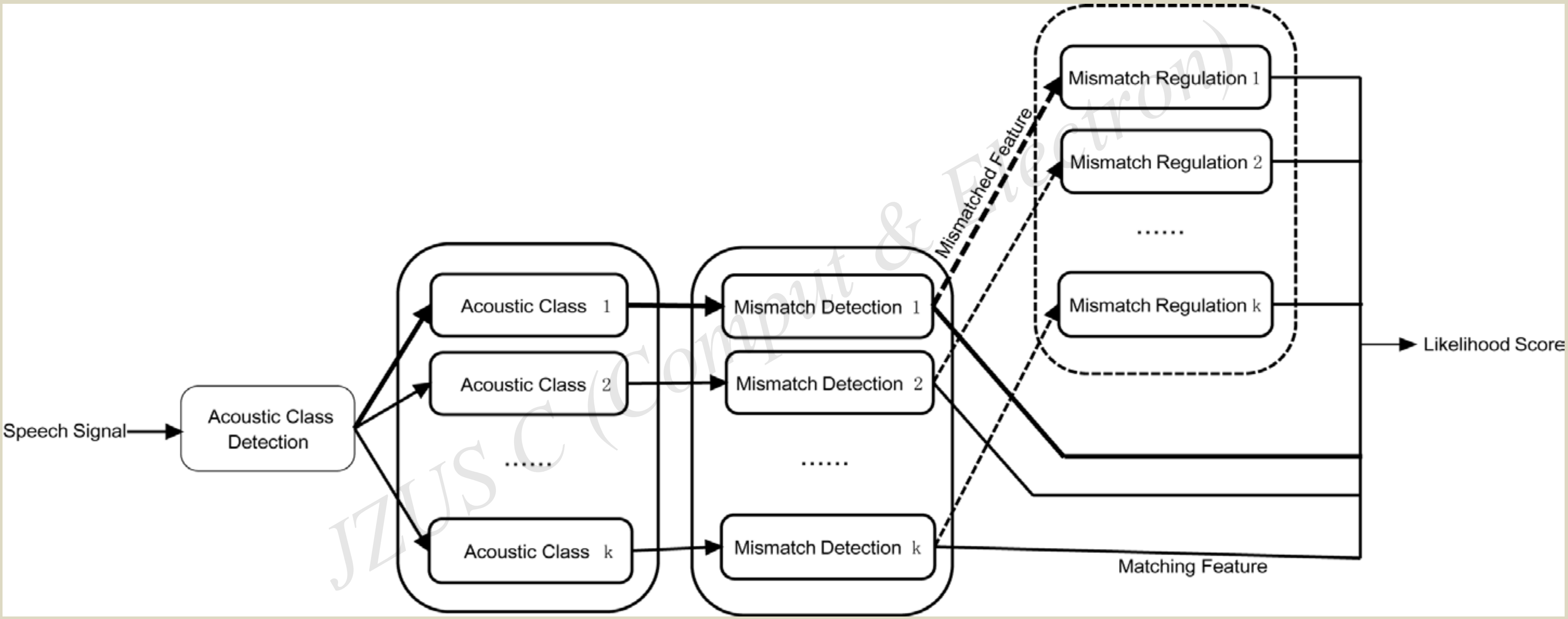
JZUS C (Comput & Electron)

Neutral-emotional discriminative power under acoustic class



The features under acoustic classes have more discriminative power than MFCC features; thus, it is more reasonable to construct the mismatched feature detector under acoustic classes.

System framework



Experiment results (I)

Table 5 Comparison between the baseline algorithm, pitch modifying method (Huang and Yang, 2008), and our feature regulation methods

Method	Identification rate (%)					
	Neutral	Anger	Elation	Panic	Sadness	Average
GMM-UBM	96.23	31.50	33.57	35.00	61.43	51.55
Pitch-mod	95.90	37.90	39.53	38.56	60.83	54.55
MDPC	94.80	38.73	41.20	39.20	60.33	54.85
MDGT	95.37	38.40	43.20	40.07	61.80	55.77
MDPGMT	95.20	41.93	44.33	44.50	65.63	**58.32**

Pitch-mod: pitch modification

JZUS-C (Comput & Electron)

Experiments on MASC (45 male, 23 female) showed that our proposed methods are superior to the baseline GMM-UBM and pitch modifying method.

Experiment results (II)

Table 8 Algorithm performance on i-vector

Method	Identification rate (%)					
	Neutral	Anger	Elation	Panic	Sadness	Average
i-vector	95.53	43.37	44.17	50.53	62.60	59.24
Feature-P	95.07	46.47	46.73	53.13	65.27	61.33
Feature-R	95.40	47.87	47.60	55.80	66.13	**62.56**

Feature-P: feature pruning; Feature-R: feature regulation

JZUS C (Comput & Electron)

Experiments on MASC (45 male, 23 female) showed that our proposed methods are superior to i-vector.

Conclusions

- We propose the mismatched feature detection method based on three sorts of acoustic classes: phoneme classes, GMM tokenizer, PGMT (probabilistic GMM tokenizer).
- Feature pruning and feature regulation methods based on (fuzzy) SVM were presented to deal with the mismatched features.
- Experiments conducted on MASC@CCNT showed that these two methods can effectively reduce the negative effects brought by emotional variability.

JZUS C (Comput & Electron)