Experiments on Emotional Speaker Recognition

Bie Fanhu
CSLT, RIIT, THU
2012-11-05

Outline

1. Introduction
2. Speaker Database Design
3. E-Norm Algorithm
4. Pitch Based Clustering Algorithm

Introduction

- When the emotion states mismatch between the training and testing speech, the capability of speaker recognition(SR) will be significantly decreased

- The emotional speaker recognition is the task to improve the robustness of SR on the emotions

Influence Factors

- Due to the speaker affected by the content of the sentence

- Due to the content of the sentence affected by the emotion of the speaker

Database Design

- Emotion Type
  - —Neutral, Happiness, Sadness, Anxiety, Anger
- Syllables Coverage
  - —21 initials and 38 finals
  - —About 500 kinds of syllables
- Content Scripts
  - —Emotional scripts
  - —Neutral scripts

E-Norm

- Norm algorithms are to decrease the mismatches between the models and the testing features
- Assumption: The scores from one speaker follow a Gaussian distribution

E-Norm

- Training Data Set
  - Train the parameters for the normalization
- Testing Data Set
  - Neutral: Including some utterances used in training data set
  - Other emotions: different from the training data set

E-Norm Result

E-Norm	Whole		Male		Female	
	EER	Threshold	EER	Threshold	EER	Threshold
Whole	19.98%	-0.8682	27.96%	-0.5685	27.25%	-0.6544
NL	5.04%	-1.5824	8.21%	-1.3874	7.23%	-1.3966
HP	21.63%	-0.7279	29.72%	-0.3886	30.21%	-0.3995
SD	21.89%	-0.8557	31.81%	-0.4137	28.22%	-0.7672
AX	22.01%	-0.7456	31.90%	-0.4548	28.71%	-0.4682
AG	26.09%	-0.7432	34.49%	-0.4659	35.77%	-0.4954

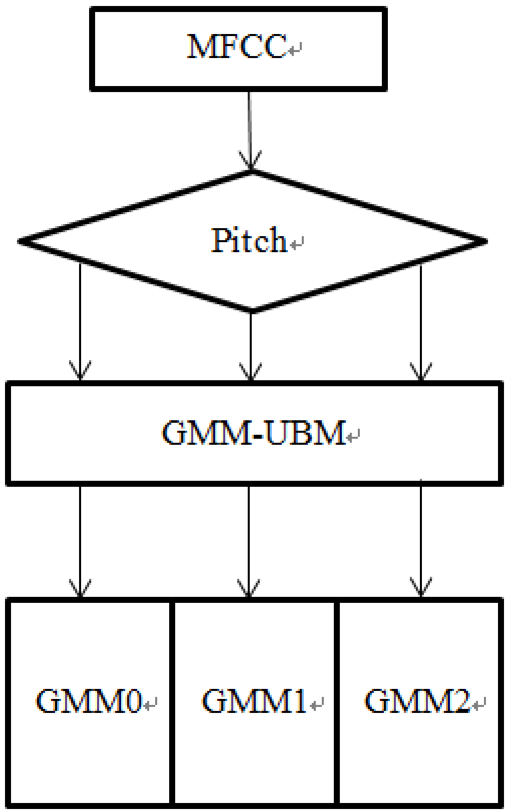
E-Norm Results

	Whole	Male	Female
Whole	20.56%	15.99%	17.60%
NL	51.86%	49.54%	36.91%
HP	10.29%	5.29%	7.84%
SD	10.32%	2.96%	4.31%
AX	6.10%	4.49%	1.91%
AG	4.78%	−1.98%	0.94%

Relative improvements

Pitch-Based Clustering

- Clustering on frames of features
- Each group training into a GMM
- 3 GMMs



Pitch Clustering

	Male(Hz)	Female(Hz)
Group1	0-120	0-120
Group2	120-196	120-290
Group3	>196	>290

Score Fusion

- Average(Score1 + Score2 + Score3)
- Max(Score1, Score2, Score3)

Results

Best	Whole		Male		Female	
	EER	TH	EER	TH	EER	TH
Whole	18.40%	0.3333	19.76%	0.311367	28.13%	0.4977
NL	14.18%	0.386433	18.81%	0.362033	21.11%	0.5098
HP	19.97%	0.3777	18.66%	0.3211	25.63%	0.6117
SD	16.56%	0.273267	19.85%	0.293967	23.56%	0.36
AX	16.37%	0.310367	16.22%	0.286033	25.93%	0.4847
AG	21.41%	0.341533	20.88%	0.299767	32.52%	0.5018

- For male, Group1 gives the best EER
- For female, Group2 gives the best EER
- The erformance of male is better than female

Thanks!