A Multiple Perception Model on Emotional Speech

Jianhua Tao Ya Li Shifeng Pan

National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences

jhtao@nlpr.ia.ac.cn

Related research

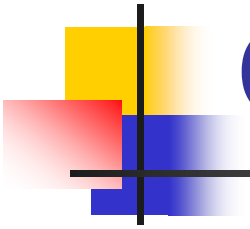
- The definition of emotion state
  some 'basic' emotion states (Happy, Sad, …)
  ……
- How to represent emotions
  Emotion vector
  ……
- How to percept the emotions in speech?

  Huber, R. *et al.*
  Dellaert, F. *et al.*
  Valery A. Petrushin
  Noam Amir
  Lee *et al.*
  Yu *et al.*

What we have done in this study?

- emotions in utterance is uncertain in most cases, To resolve the ambiguity of emotion perception, array of perception experiments were conducted.

Corpus and labeling



- 5.5 hours, 500 sentences and read by 4 professional speakers with 5 emotions.
- segmentally and prosodically annotated with break index, F0 values are extracted and manually checked .
- Training set : test set 8:2

Perception experiment

- The 15 subjects were asked to note the emotion they perceived with one or two emotion states from a list of “happiness, fear, sadness, anger and neutral.”

- Sample

Utt No	One choice	Two choice	
1	Sad	Sad	Fear

Perception experiment

Table 1: Perception experiment of ‘Happiness’ utterances (%)

		h	f	s	a	n
one choice		77.8	0.4	0.2	0.6	5.6
two choices	first choice	10.9	0.6	0.7	0.4	3.0
	second choice	3.1	0.7	0.4	0.4	10.9

Table 2: Perception experiment of ‘Fear’ utterances (%)

		h	f	s	a	n
one choice		0.0	79.7	0.7	0.0	0.7
two choices	first choice	0.0	15.6	2.4	0.2	0.7
	second choice	0.0	3.1	12.8	0.9	2.1

Perception experiment

Table 3: Perception experiment of ‘Sadness’ utterances (%)

		h	f	s	a	n
one choice		0.5	11.3	47.2	0.9	4.0
two choices	first choice	0.0	11.1	17.7	1.7	5.6
	second choice	0.5	10.4	12.5	4.3	8.3

Table 4: Perception experiment of ‘Angry’ utterances (%)

		h	f	s	a	n
one choice		0.5	11.3	47.2	0.9	4.0
two choices	first choice	4.5	0.0	0.0	9.5	1.4
	second choice	4.0	0.0	0.2	4.9	6.4

Perception experiment

Table 5: Perception experiment of ‘Neutral’ utterances (%)

		h	f	s	a	n
one choice		0.9	3.6	0.0	0.5	83.9
two choices	first choice	0.5	1.2	0.0	0.3	9.0
	second choice	3.6	2.6	1.2	2.6	1.0

Though the professional speakers were asked to read every sentence with specific emotion, perception experiment shows the listener can make sure his/her decision, which is not exactly as the same as the speaker performed

Perception experiment

- The emotion vector of a utterance is defined as:

$$E = (n, h, s, a, f)$$

- Here:

$$n = \frac{1}{N}\sum_{i=1}^{N}\omega_{n,i} \qquad h = \frac{1}{N}\sum_{i=1}^{N}\omega_{h,i} \qquad s = \frac{1}{N}\sum_{i=1}^{N}\omega_{s,i}$$

$$a = \frac{1}{N}\sum_{i=1}^{N}\omega_{a,i} \qquad f = \frac{1}{N}\sum_{i=1}^{N}\omega_{f,i}$$

- For “one choice” result, $w = 1.0$
- For “first choice”, $w = \alpha \ (\alpha < 1)$
- the “second choice”, $w = 1 - \alpha$

Perception model

- Classification and regression model
- Features
- One tree VS. five trees
- Acoustic influence analysis

Perception model

- Features:

Duration,

f0-range,

f0-variation,

f0-maximum,

f0-minimum,

f0-mean,

position of f0 peak in the utterance,

position of f0-minimum in the utterance,

power-range,

power variation,

power-maximum,

power-minimum,

power-mean,

position of power-peak in the utterance,

position of the power minimum in the utterance

Voice quality (LF, Ee, Ra, Rk,Rg, Oq)

Perception model

- One tree for all
  precision 72% with 36 leaf nodes

  the results matches well with the majority of listener responses, however, it does not explain the differences in the opinions of the individual respondents when they listened to the same or equivalent samples.

  how to represent this probabilities?

Perception model

- Five trees

  - Each tree was trained with the same acoustic data for the independent variables, and with the probabilities of a response in its own category as the dependent variable.

  - “final” emotion state decision

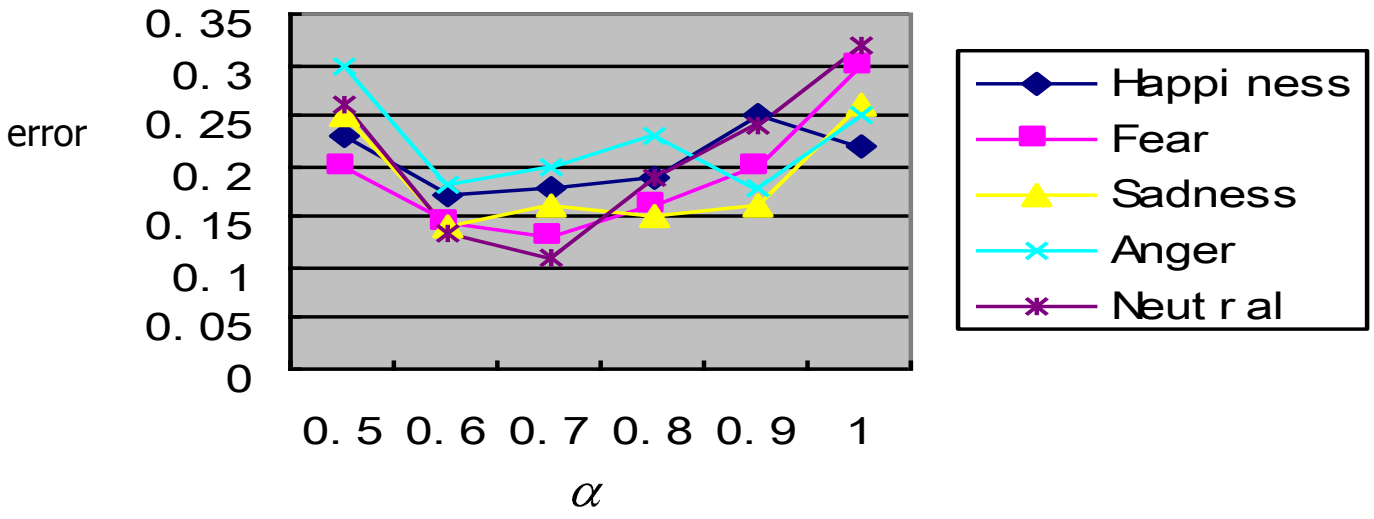
- Results to five trees

Table 6: Outputs of the models

$$\alpha = 0.6$$

	Happiness	Fear	Sadness	Anger	Neutral
Utt1	0.4879	0.0181	0.0264	0.3437	0.1238
Utt2	0.5103	0.0189	0.0276	0.3595	0.0837
…					

Fig 1 The distribution of emotion vector prediction errors from $\alpha = 0.5$ to $\alpha = 1.0$



Acoustic influence analysis

The contribution toward the importance of a predictor appears in the n'th surrogate is defined as:

importance_contribution_node_i

*= (p ^ n) * improvement*

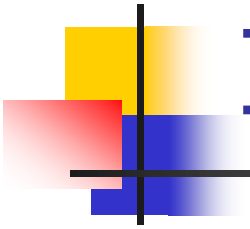
p is the "surrogate improvement weight"

Table 7: Ranking score of input parameters

Parameters	Ranking Score
F0 mean	100.00
F0-maximum	98.79
F0-range	98.64
Ee	57.63
Duration mean	48.34
Duration Range	38.78
Position of F0 minimum	36.36
Power	34.86
Rd	33.52
……	……

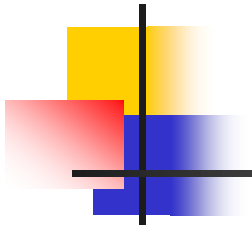
Conclusion

- different people are sensitive to different facets of information, especially the emotion.
- It may be more appropriate to use a vector of emotions to represent this uncertainty.
- five decision tree classifiers is better than one in this task.
- The results show important cues on how to do multiple perception of the emotional speech.



In future

- try more efficient features to predict the emotion(s) in speech
- carry out this experiment on more realistic speech data



Thank you!