The role of F0 trajectory in the emotion identification

윤태진
성신여자대학교 영어영문학과
Two emotional distinction theories

• The discrete emotion theory
  • Basic discrete emotions exist:
    (1) surprise, (2) interest, (3) joy, (4) rage (5) fear (6) disgust (7) shame (8) anguish
  • Individual emotions have biological and neurological profiles

• The dimensional theory
  • Two emotional dimensional spaces distinguish emotions
    (1) valence – how positive or negative an emotion is
    (2) arousal – the intensity of an emotion
The discrete emotion approach

- Emotions are discrete, measurable, and physiologically distinct.
- Certain emotions appeared to be universally recognized.

→ Many studies have examined the vocal characteristics of speech in hope of defining a vocal signature for each basic emotion (Russell 2003)
The Dimensional approach

- The strongest single association found for vocal acoustic have been with the sender’s general arousal level.
- High-arousal emotions such as anger and joy have similar characteristics low arousal emotions such as sadness
  - greater loudness,
  - higher pitch
  - faster speech
- Few works have concentrated on distinguishing emotions between positive- and negative-valence emotions such as anger and joy.

Research topic

- F0 contours contains discriminatory information about emotions.
- Very few can be found in the literature that made the efforts to describe the shape of f0 contours directly in classifying emotions.
The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)

- The RAVDESS dataset is a multimodal validated English dataset that contains speech, song, and video files that represent 8 emotions.
- The portion of the dataset that I use in this study is the speech audio files that are represented by 1440 wave file.
- Twenty-four professional actors (12 female and 12 male) with 60 trials for each actor produced the 1440 wave files (24×60 =1440).
The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)

- The actors vocalized two sentences in a neutral North American accent.
  - “Kids are talking by the door”
  - “Dogs are sitting by the door”
- The emotions
  - neutral, calm, happy, sad, angry, fearful, surprise, and disgust
- Each expression is produced at two levels of emotional intensity (normal and strong) except for the neutral emotion that is recorded in a normal intensity only.
Generalized Additive Mixed Modeling

- In Linear Model, the mean of data is modeled as a sum of linear terms
  \[ y_i = \beta_0 + \sum_j \beta x_{ji} + \varepsilon_i \]

- In Generalized Additive Mixed Model, the mean of data is modeled as a sum of \textit{smooth} functions (= smooths)
  \[ y_i = \beta_0 + \sum_j s_j(x_{ji}) + \varepsilon_i \]

GAMM approach to the F0 contour modeling
Gamm Modeling

Formula:
F0 ~ Emotions + s(Time, by = Emotions, k = 10) + s(Actor, bs = "re") + s(Actor, Emotions, bs = "re")

Parametric coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 7.9017 | 1.0637 | 7.429 | 1.10e-13 *** |
| Emotionsangry | 6.9618 | 0.6277 | 11.091 | < 2e-16 *** |
| Emotionscalm | -0.6860 | 0.6277 | -1.093 | 0.274459 |
| Emotionsdisgust | 2.2108 | 0.6277 | 3.522 | 0.000428 *** |
| Emotionsfear | 7.3391 | 0.6277 | 11.692 | < 2e-16 *** |
| Emotionshappy | 5.7515 | 0.6277 | 9.163 | < 2e-16 *** |
| Emotionssad | 2.8123 | 0.6277 | 4.480 | 7.46e-06 *** |
| Emotionssurprise | 6.2753 | 0.6277 | 9.997 | < 2e-16 *** |

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) = 0.649  Deviance explained = 64.9%
fREML = 1.0445e+06  Scale est. = 20.228  n = 357120
Pair-wise comparison of contours
Conclusion

- I attempted to model Emotions using F0 contours as an input to generalized additive model (GAM).
- The present approach has predictive power (64.9%).
- The additive model provides visualized aids and makes us better understand validity data obtained from human labelers.
THANK YOU