

## A study of emotional articulation in the framework of the Converter/Distributor model

Jangwon Kim, Donna Erickson, Sungbok  
Lee, Shrikanth S. Narayanan



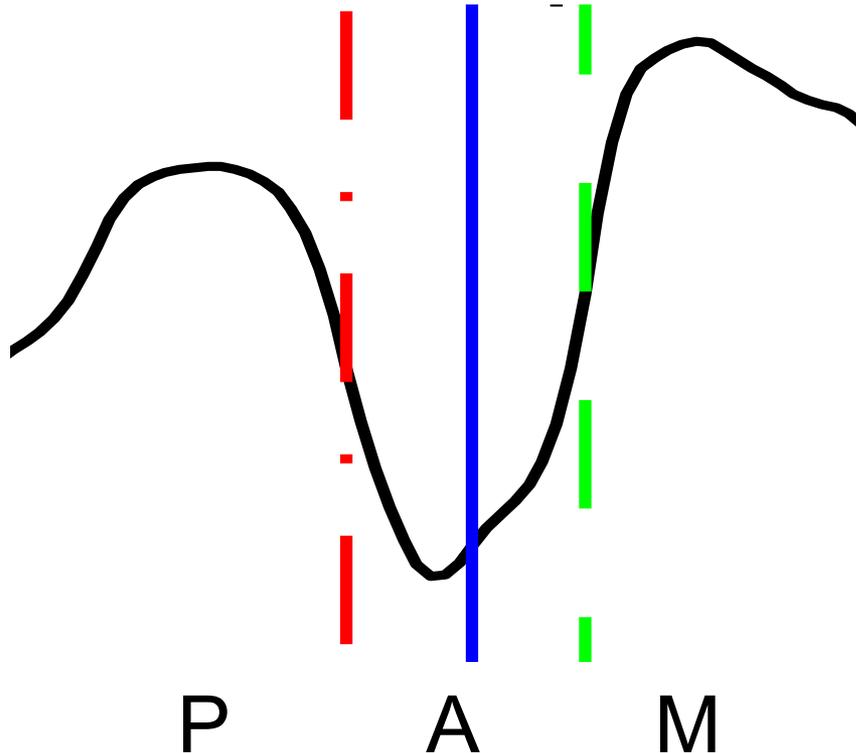
- Goals
- What is the C/D model? (relevant parts only)
- Why do we use this model?
- Dataset
- Analysis results
  - invariant properties in the C/D model for emotional speech
  - Variation patterns with emotion
- Discussion and future works

To investigate the invariant properties and variation patterns of articulatory movements in emotional speech from the perspective of the Converter/Distributor (C/D) model\*.

- Aim at systematic description and understanding of articulatory control mechanism associated with emotion using the C/D model
- Test the validity of the assumptions in the C/D model using emotional speech data
- Provide practical methodology for articulatory modeling and (re-)synthesis with emotion coloring.

\*O. Fujimura, "The C/D model and prosodic control of articulatory behavior," *Phonetica*, vol. 57, no. 2-4, pp. 128–138, 2000.

- Iceberg point for each demisyllable is determined on the trajectory of the critical articulator for the consonant



--- : Iceberg time point for onset

— : syllable center point

--- : iceberg time point for coda

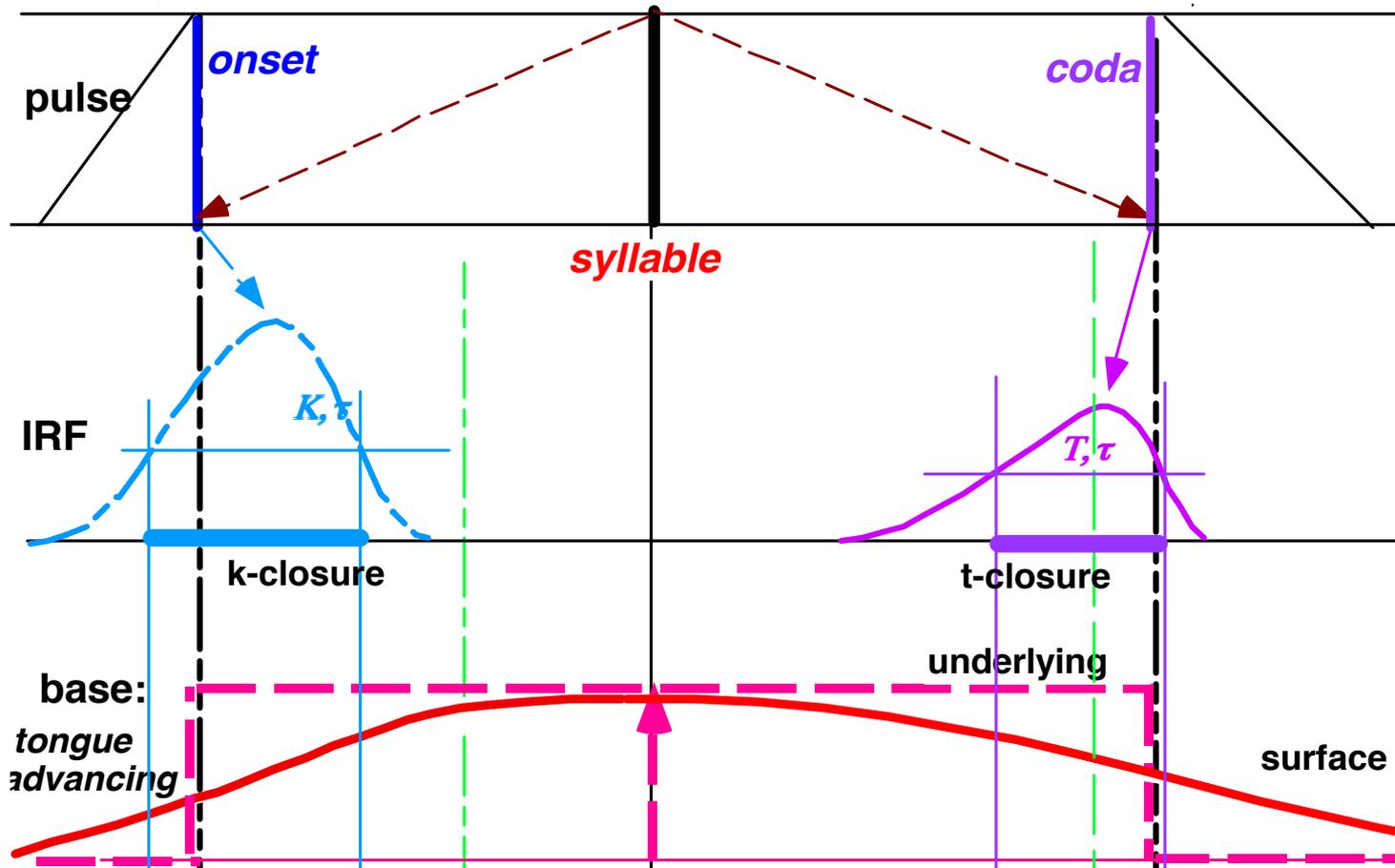
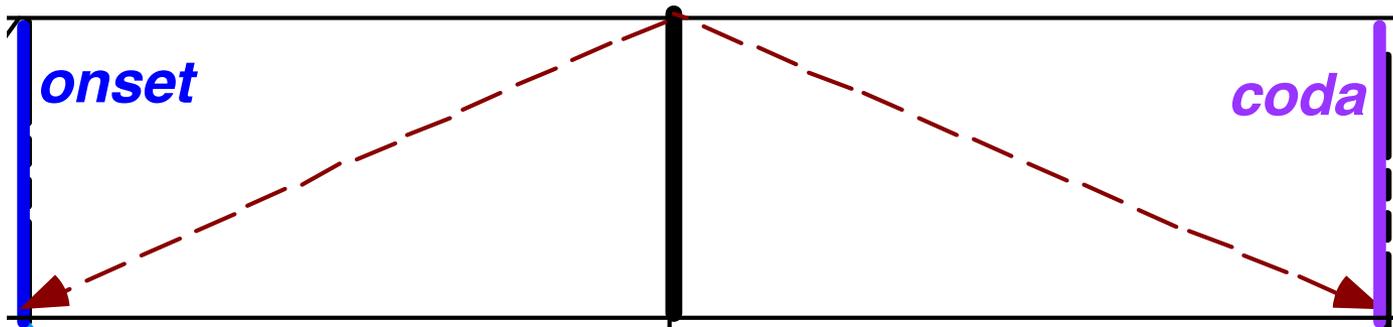


Fig2: 'kit' (CD diagram)\*

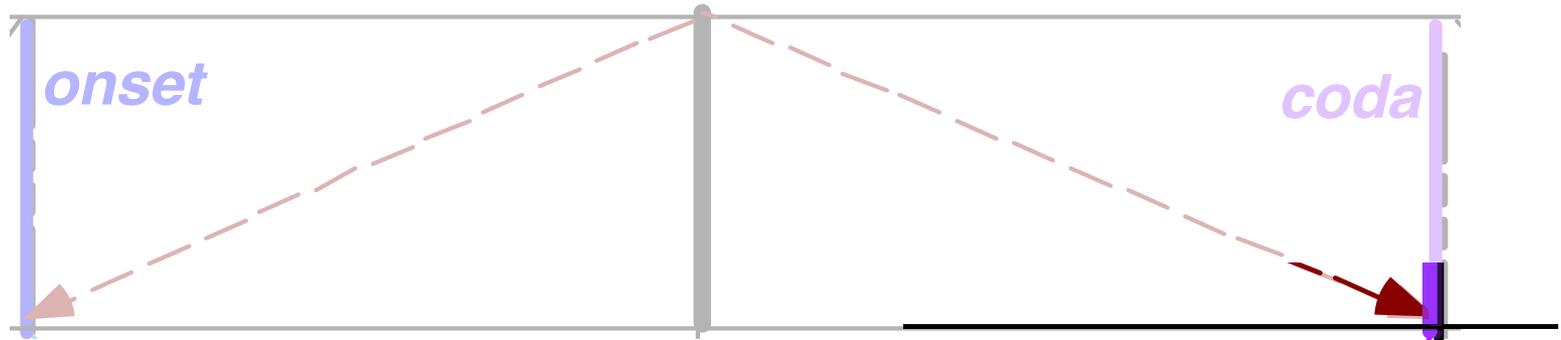
\*O. Fujimura, "The C/D model and prosodic control of articulatory behavior," *Phonetica*, vol. 57, no. 2-4, pp. 128–138, 2000.



*Top panel of the Fig. 2*

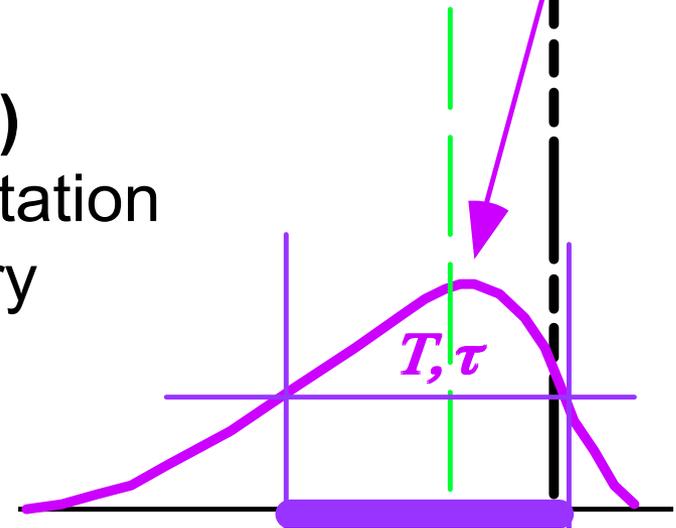
### **Pulses and syllable triangles:**

- High-level representation of metrical organization
- Prominence and (abstract) articulatory duration for each syllable



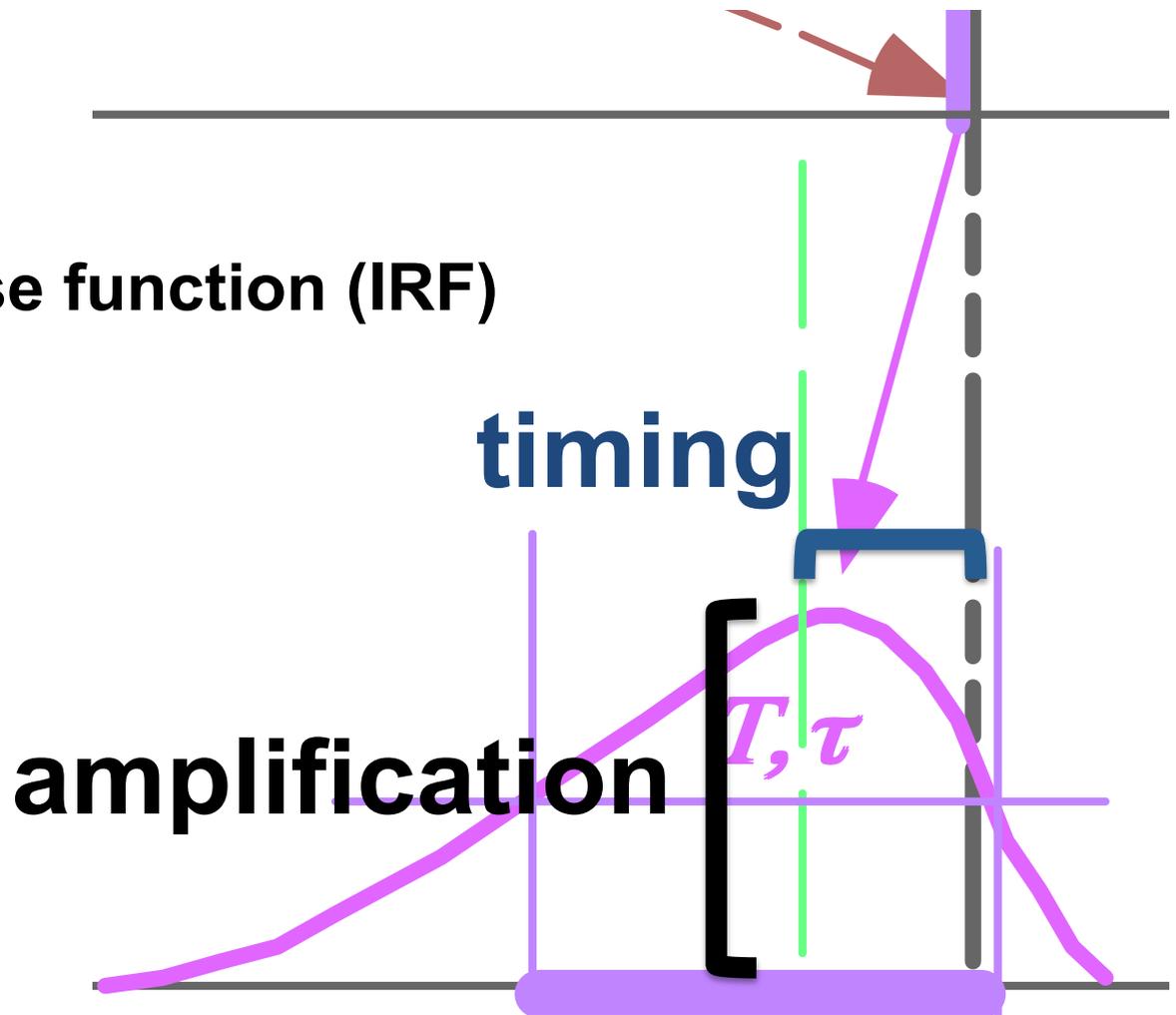
## Impulse response function (IRF)

- Relates the high-level representation and the surface-level articulatory movements.



*Middle panel of the Fig. 2*

Impulse response function (IRF)



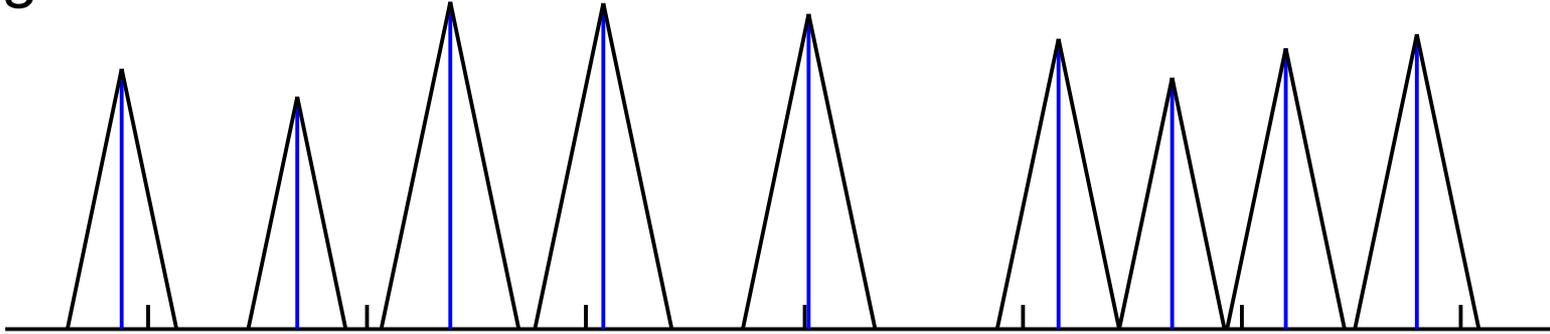
Middle panel of the Fig. 2

By studying the C/D model parameters of emotional speech, we can understand...

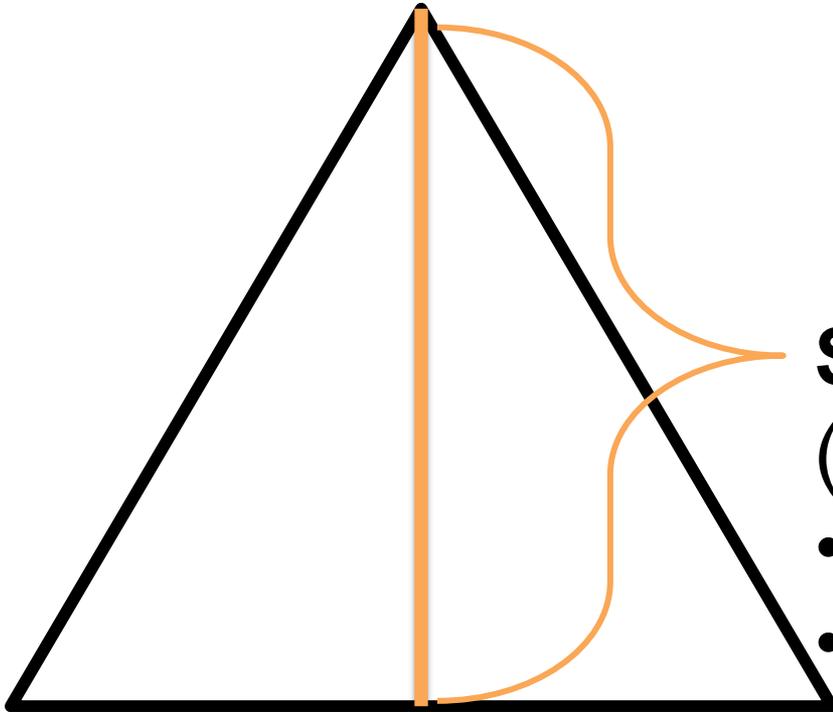
- High-level temporal organization of articulatory movements.
- Important surface-level articulatory variability for emotional speech (movement speed, range, timing, etc.)
- Relationship between the high-level representation and the surface-level articulatory movements.

## Temporal organization in syllable level

e.g. Pam said bat that fat cat at that mat

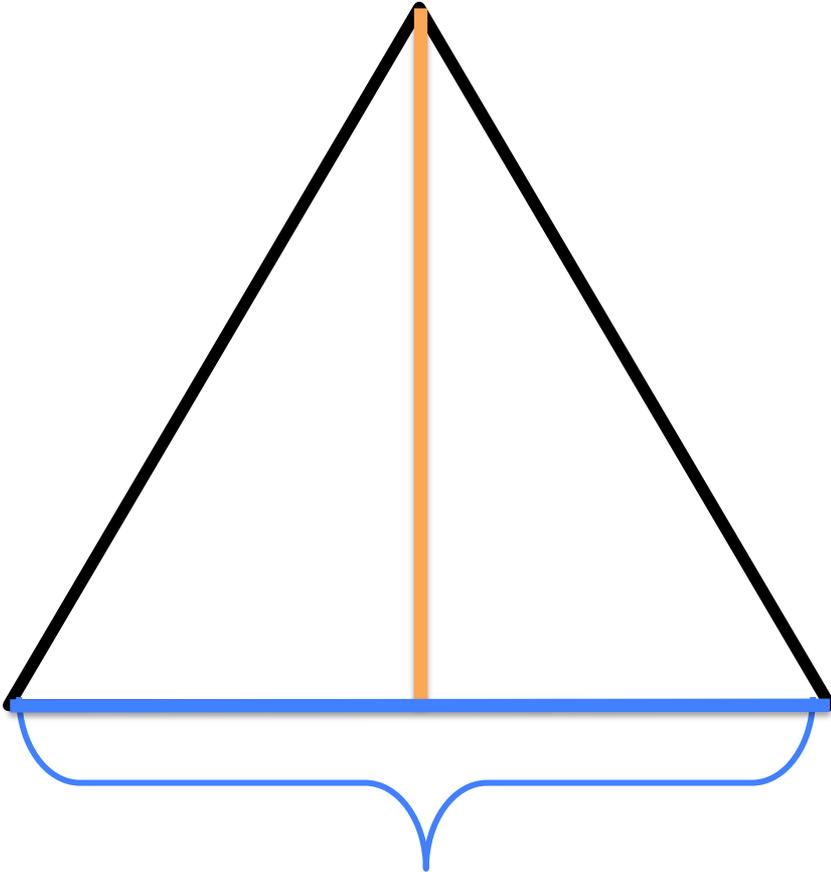


- Rhythmic pattern of consecutive syllables
- Syllable prominence
- *Abstract* articulatory duration for each syllable



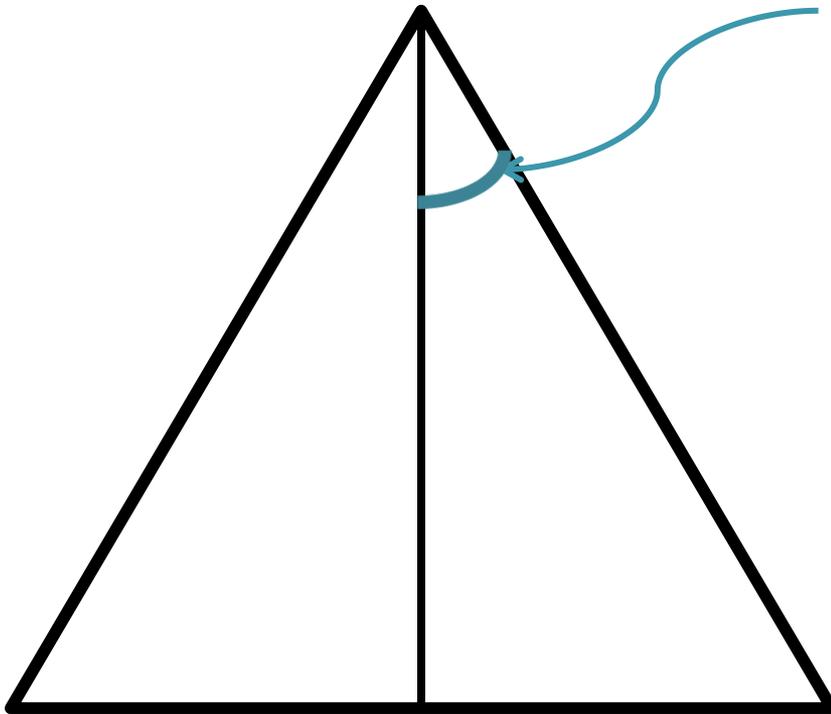
## Syllable Magnitude (triangle height)

- Jaw excursion
- Syllable prominence



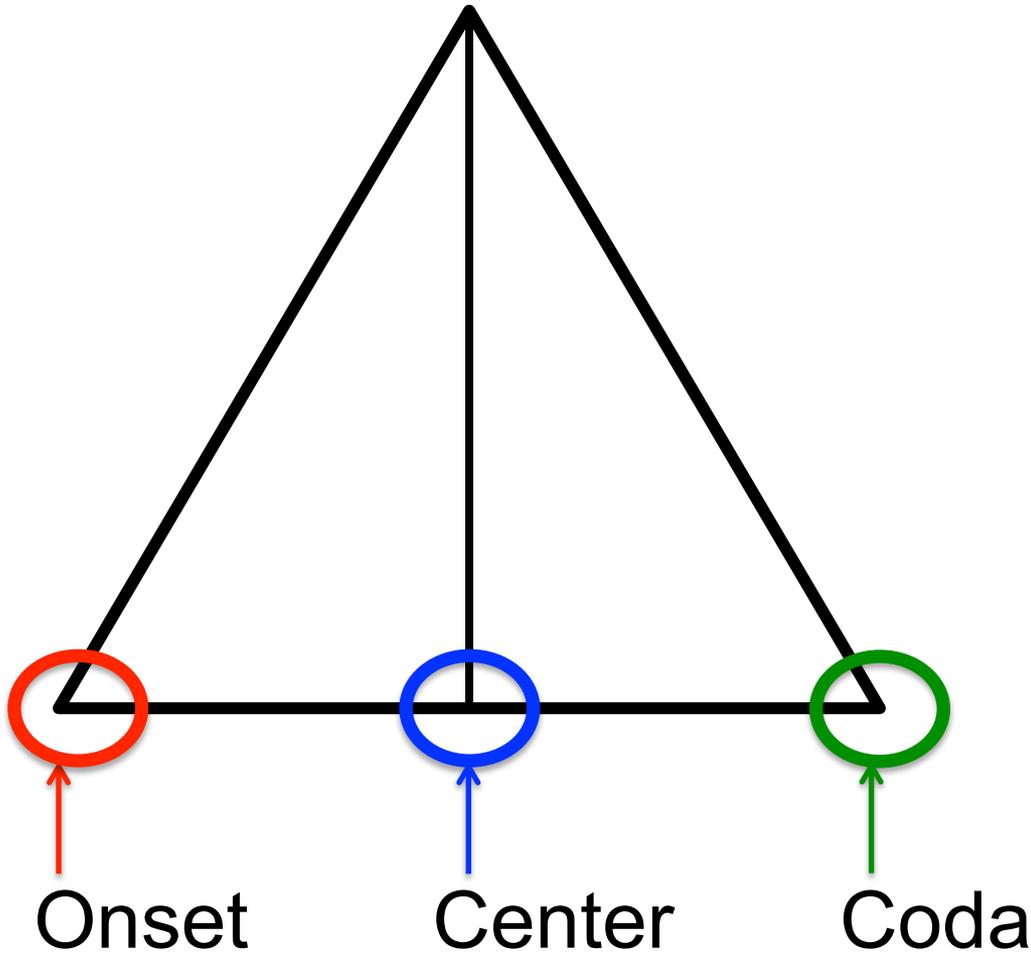
**Syllable duration** (triangle width)

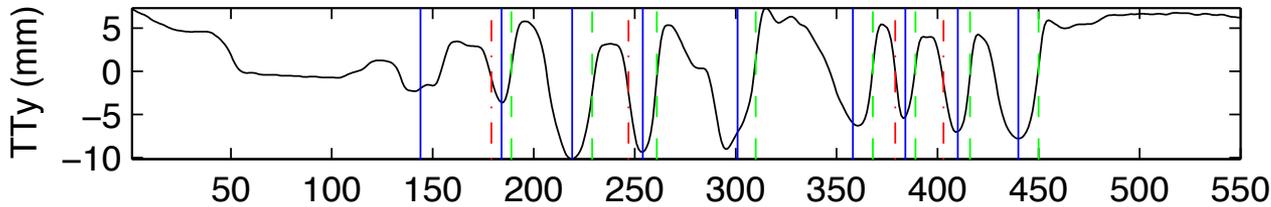
- driven from articulatory movement



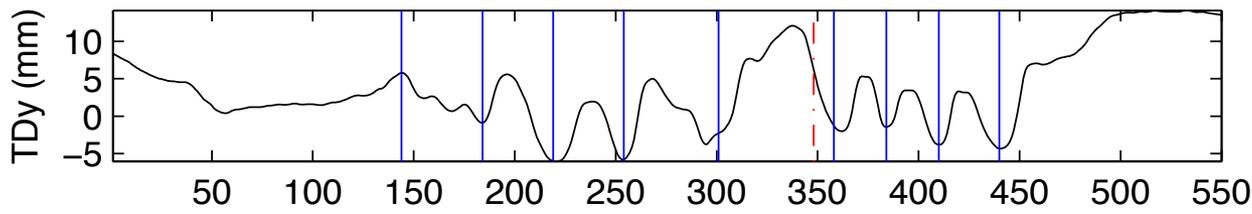
## Shadow angle

- Identical across triangles

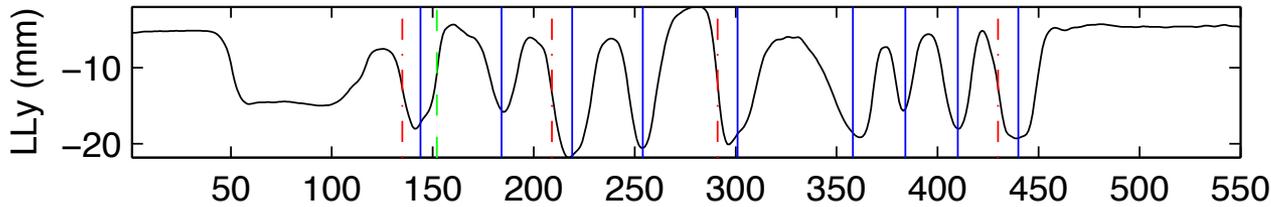




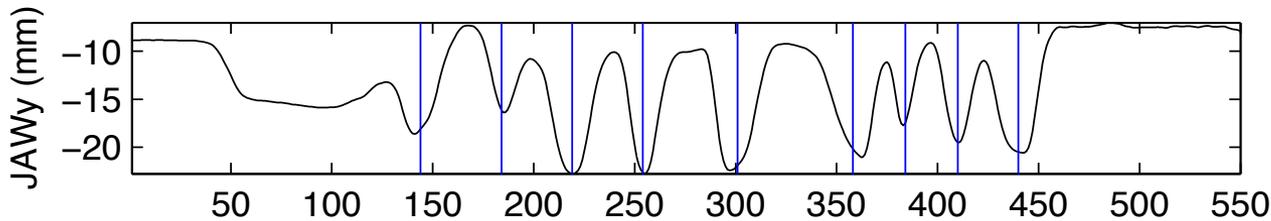
--- : Iceberg time point for onset



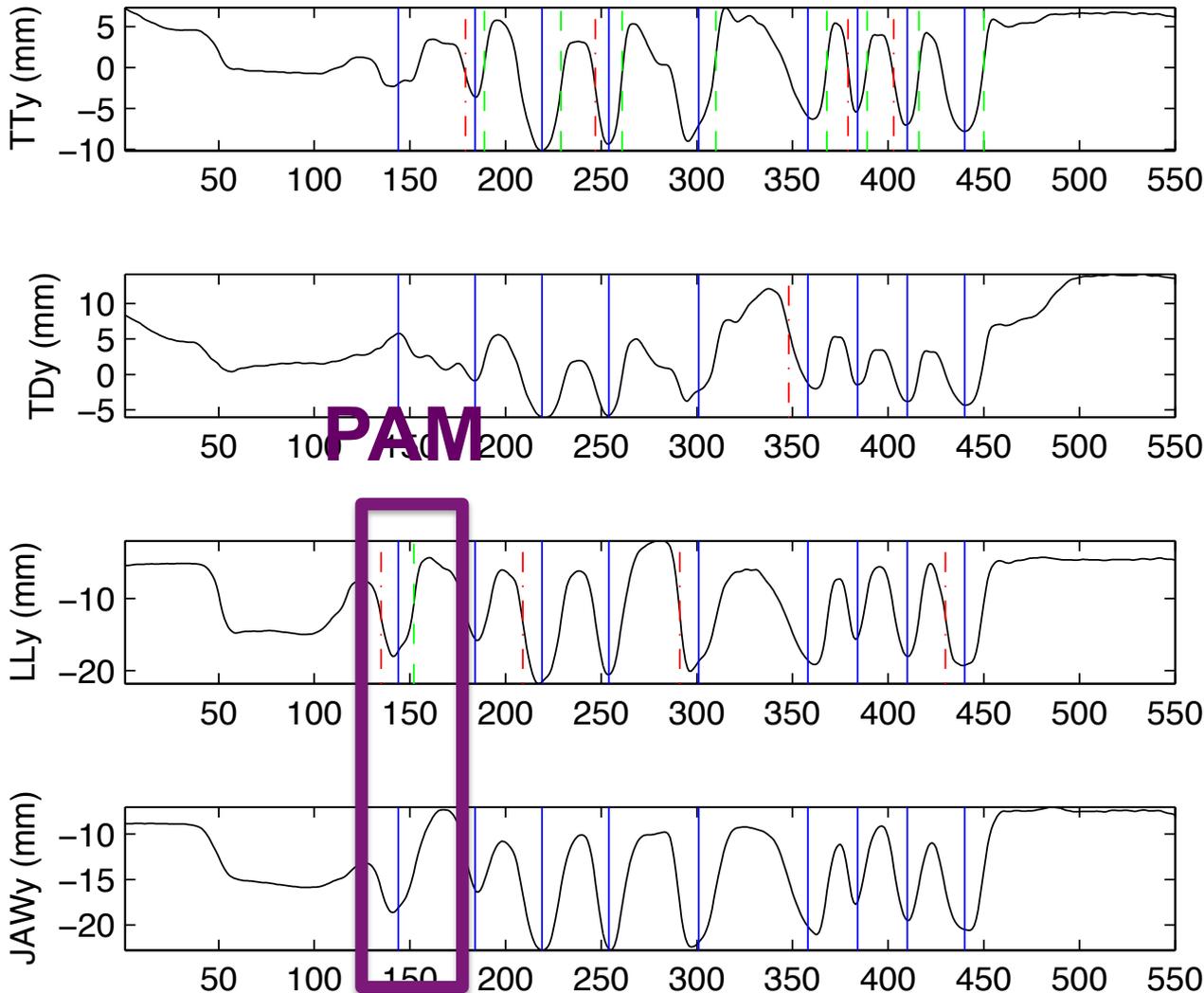
— : syllable center point



--- : Iceberg time point for coda



**Sentence: Pam said bat that fat cat at that mat**



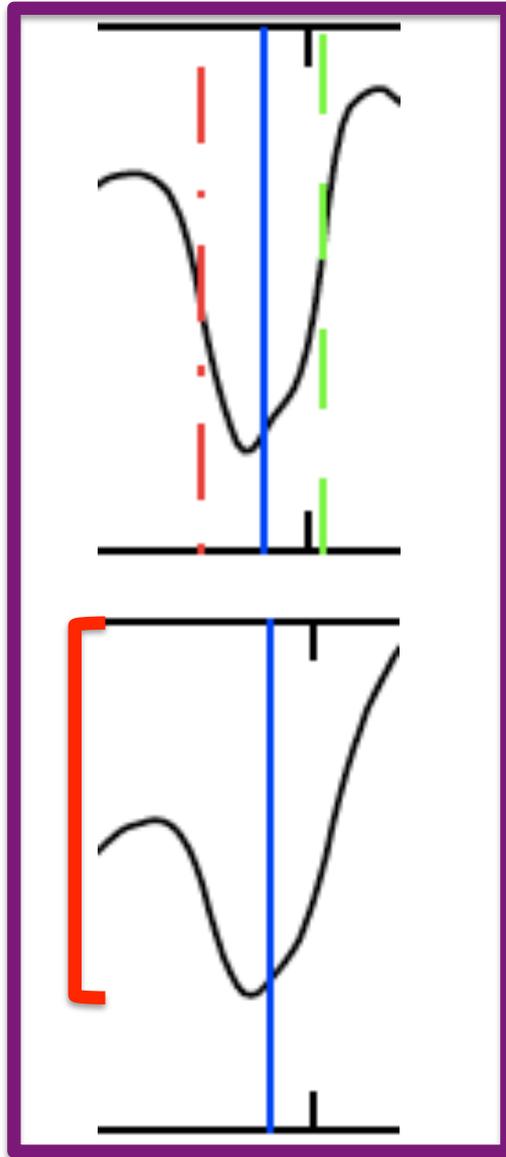
--- : Iceberg time point for onset

— : syllable center point

--- : iceberg time point for coda

**Sentence: Pam said bat that fat cat at that mat**

LL\_y



--- : iceberg time point for onset

— : syllable center point

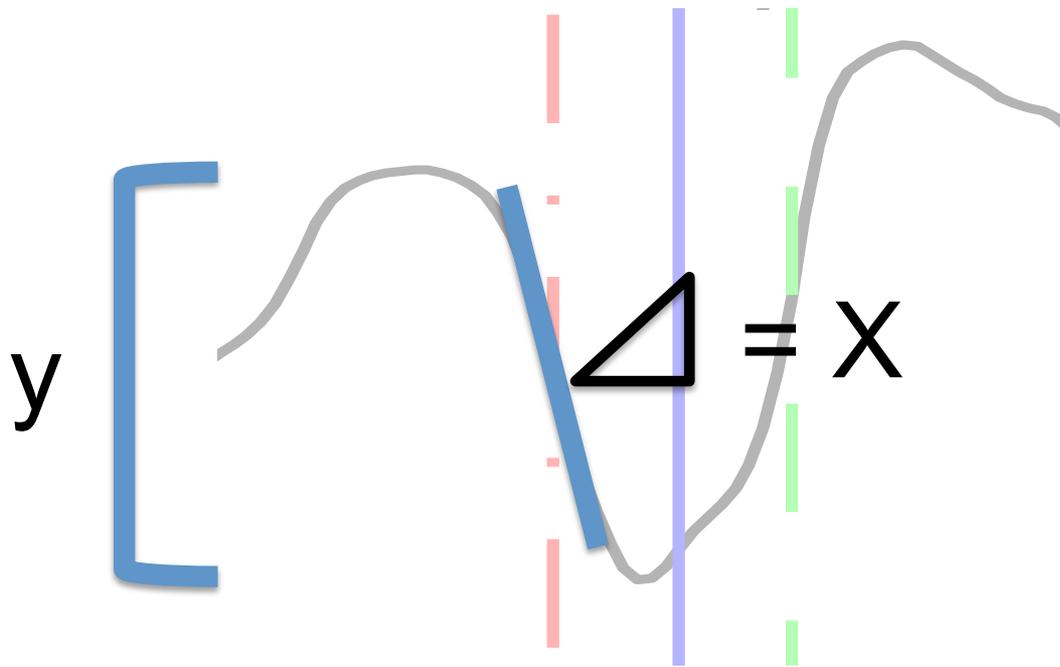
--- : iceberg time point for coda

Jaw\_y

# PAM

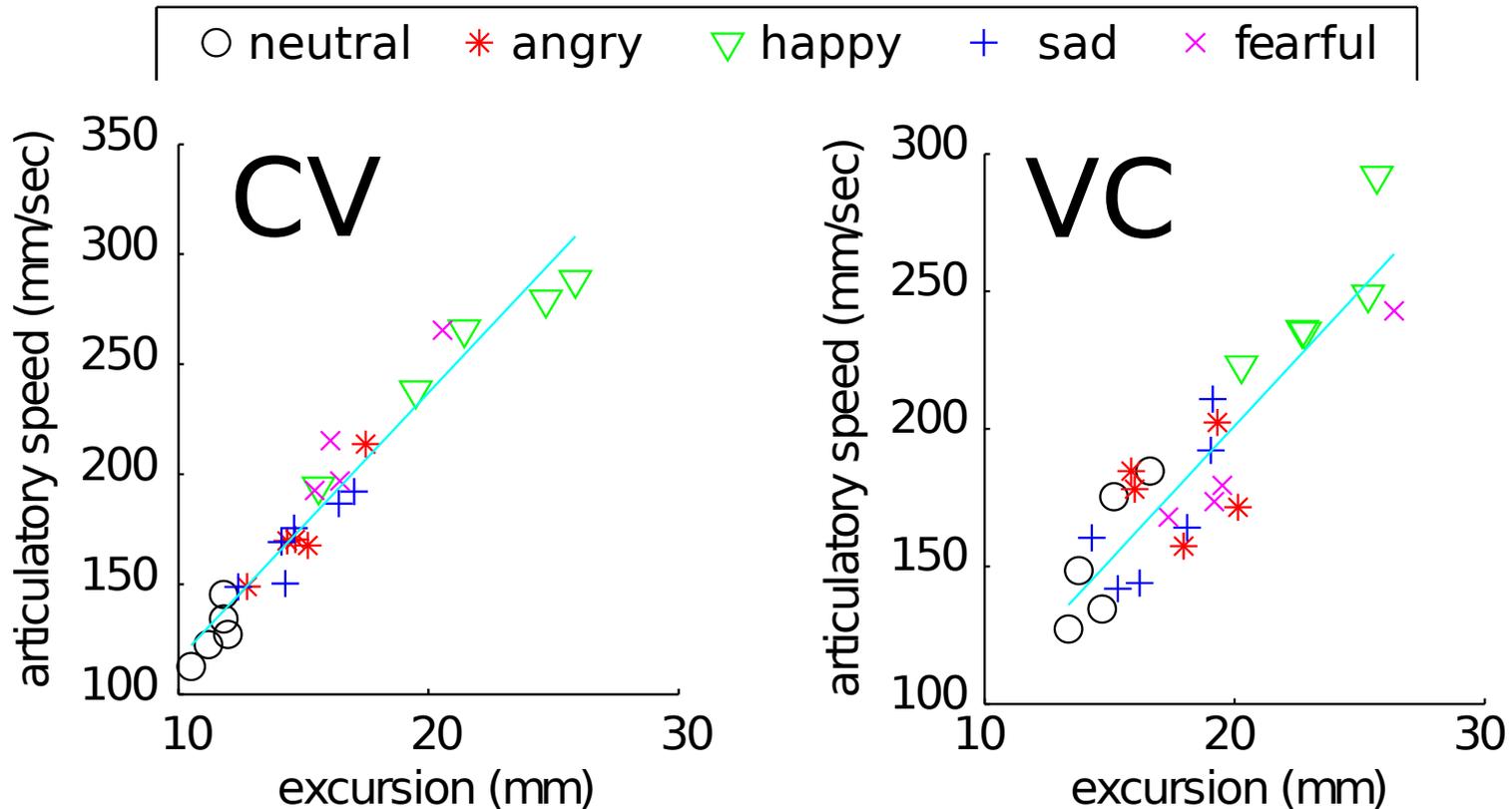
- ElectroMagnetic Articulography (EMA) dataset
  - the NDI WAVE system
  - Tongue tip, tongue dorsum, lower lip, jaw are analyzed.
  - Interpolation, smoothing, occlusal plane correction
- A female native speaker of American English
- “Pam said bat that fat cat at that mat” 
  - 5 emotions: neutral, anger, happy, sad, fear
  - 5 repetitions
  - Syllables: CVC
- Emotion quality evaluation by 11 native speakers of American English

- Linear relationship between speed at the iceberg point and the movement range of critical articulator

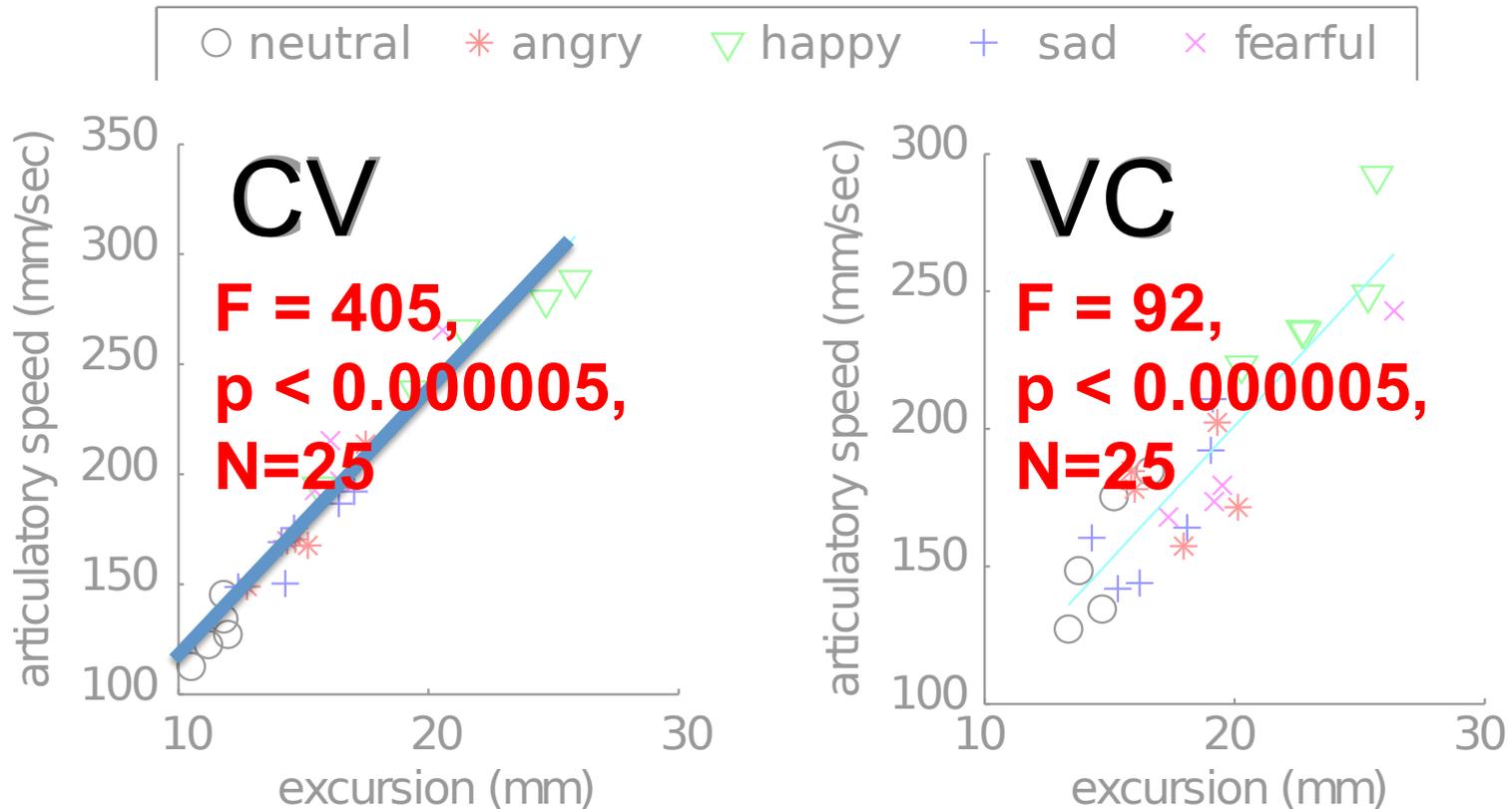


$$y = C x, C \text{ is constant}$$

- The articulatory speed and excursion of critical articulator are strongly linear in emotional speech.

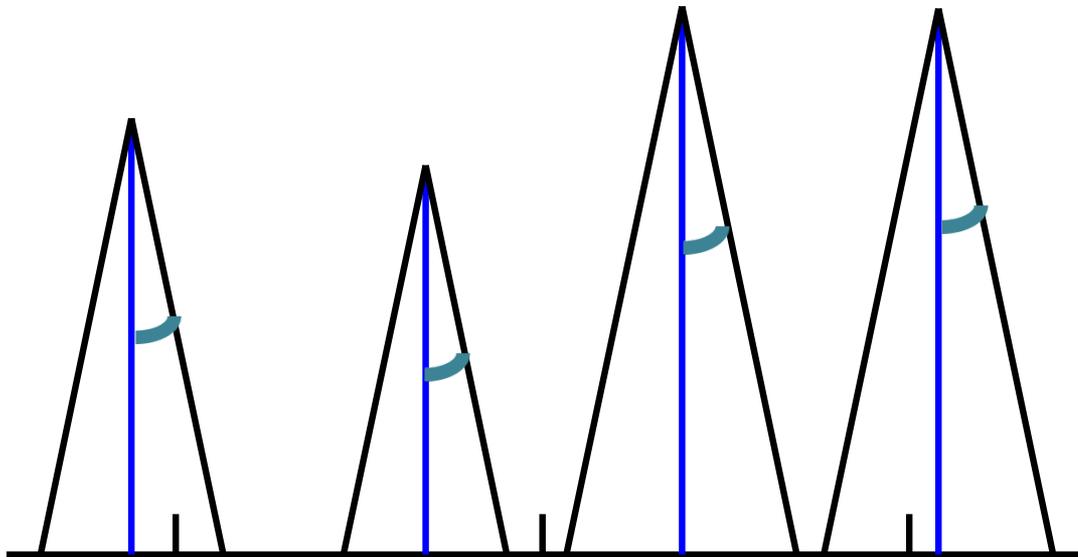


- The articulatory speed and excursion of critical articulator are strongly linear in emotional speech.



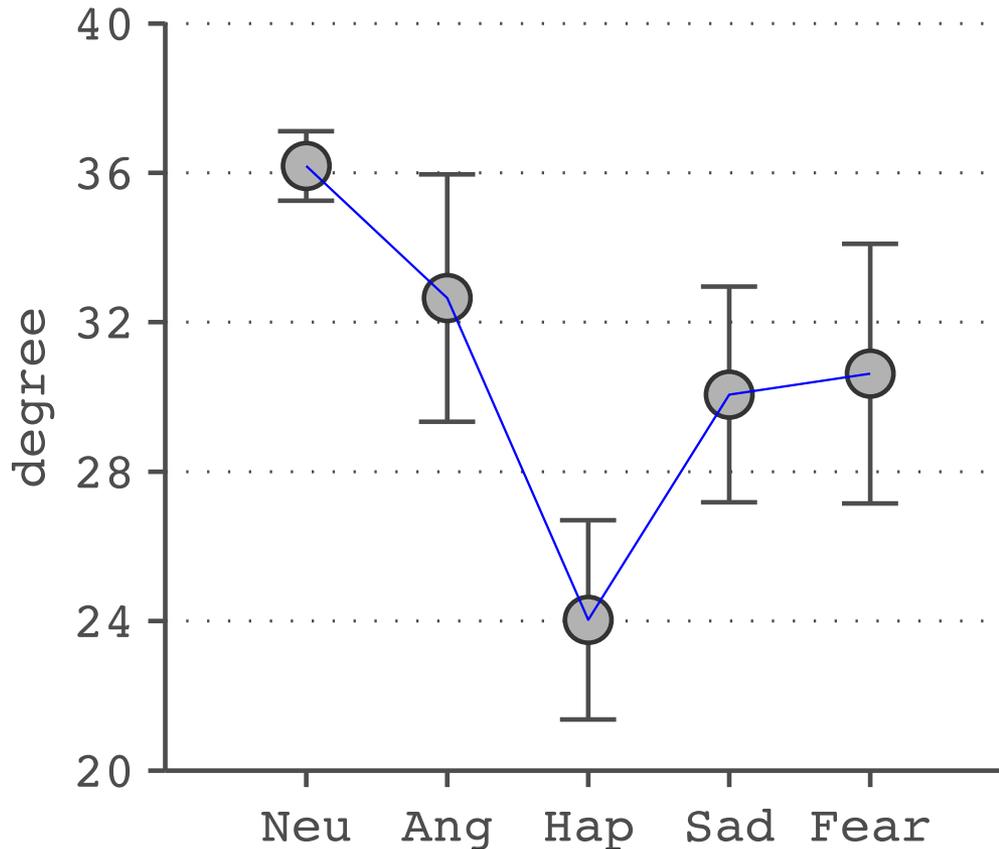
*Excursion of the crucial articulator (for /P/ & /M/) and the articulatory speed at icebergs in CV/VC demisyllables for “PAM.”*

- Identical shadow angle, regardless of stress conditions of the syllables.

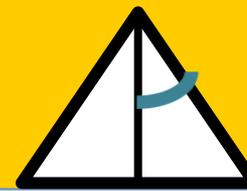




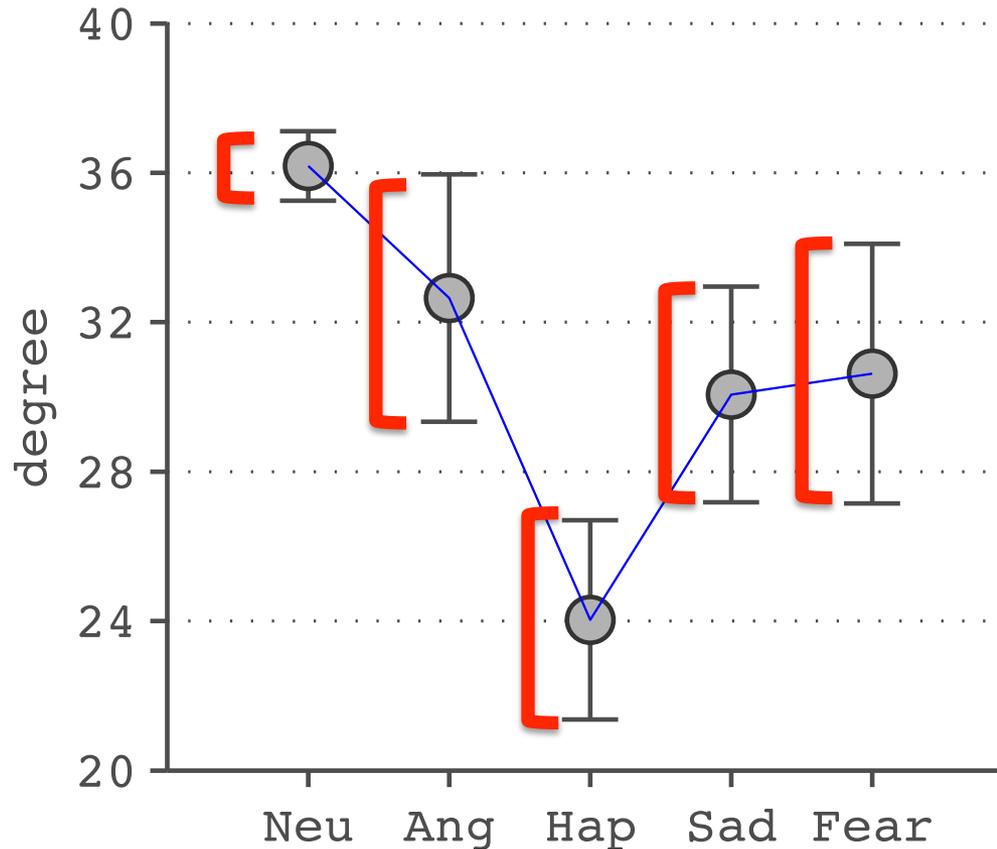
- Emotion affects shadow angle, thus not invariant.



*Errorbar plot of the shadow angles for each emotion.*



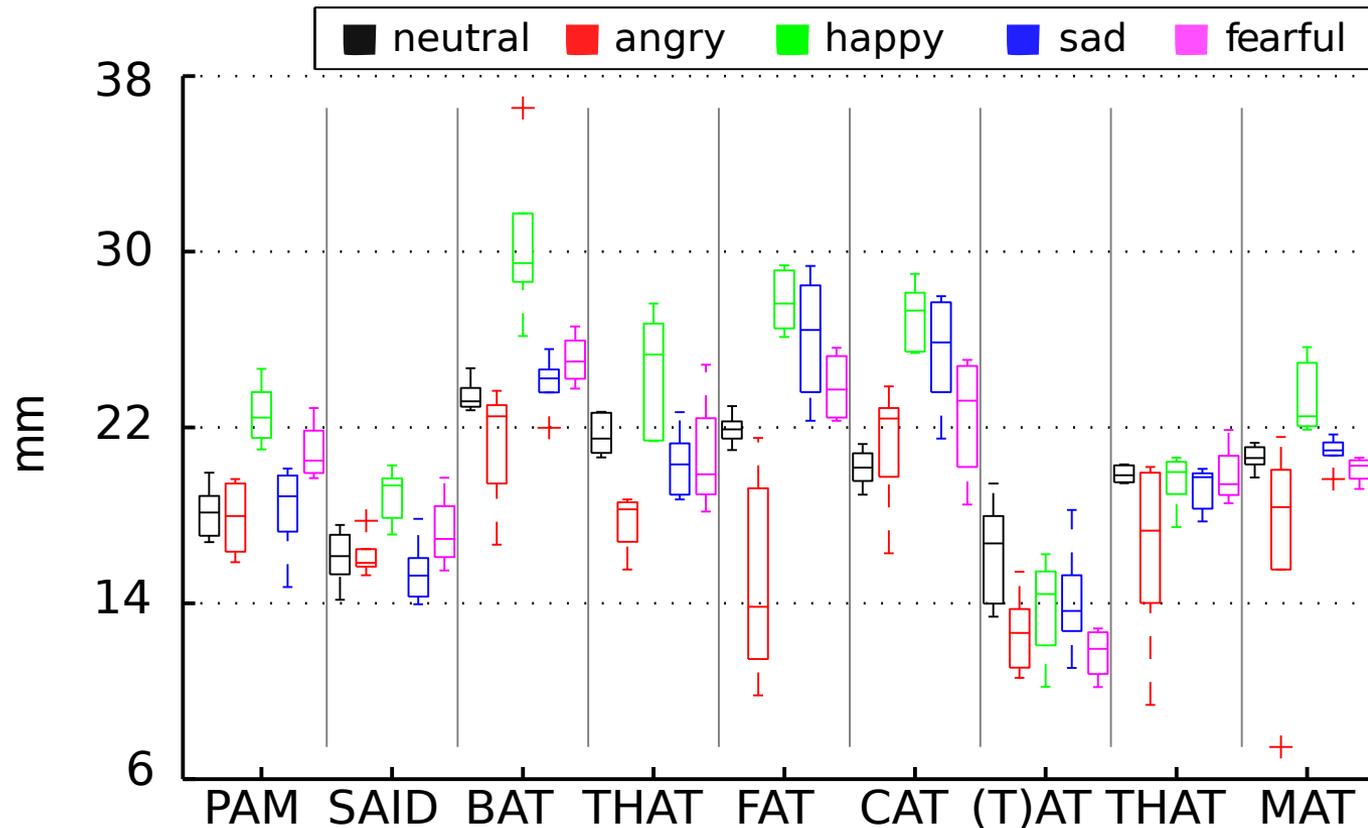
- Emotion affects shadow angle, thus not invariant.



*Errorbar plot of the shadow angles for each emotion.*



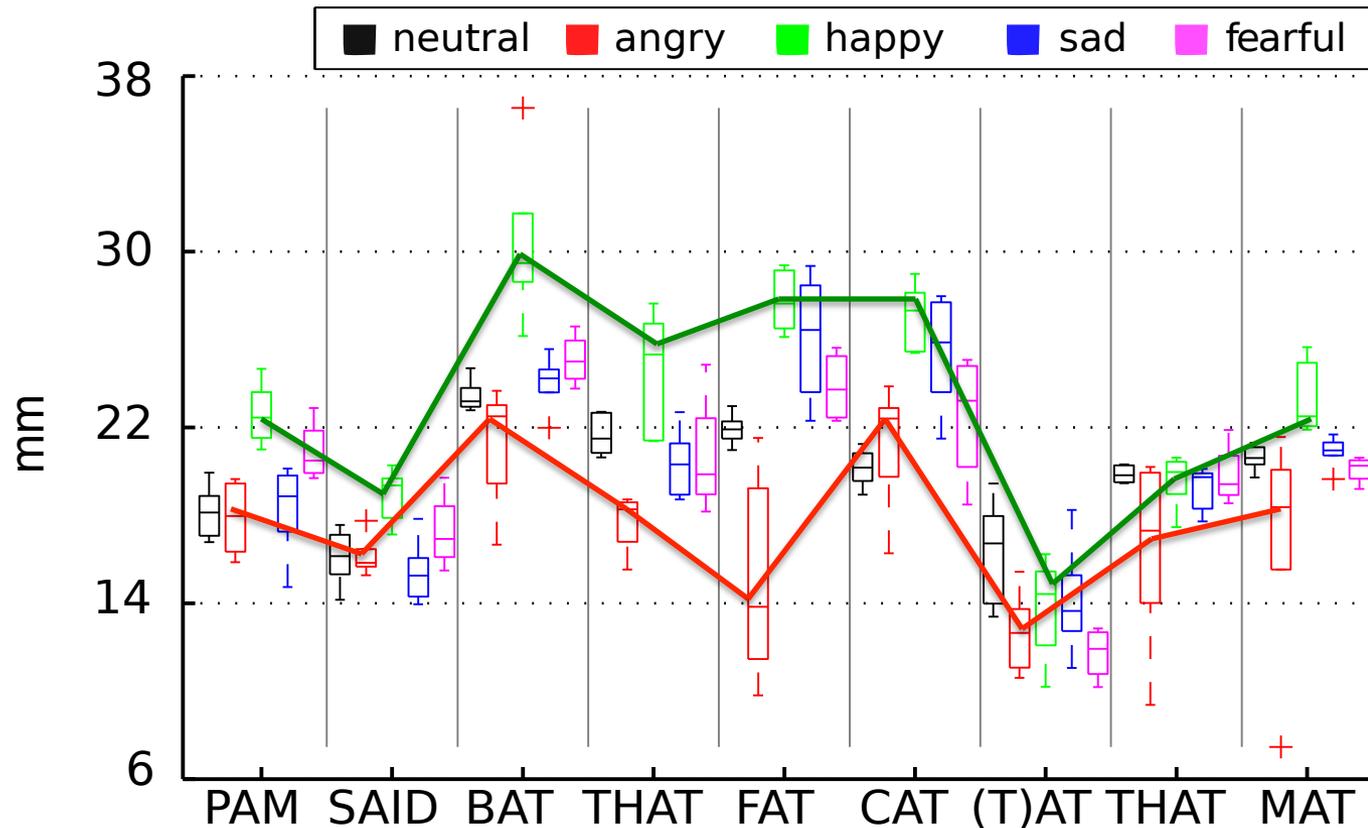
- Emotion affects syllable magnitude systematically



*Errorbar of syllable magnitude for each syllable and each emotion in the utterance*



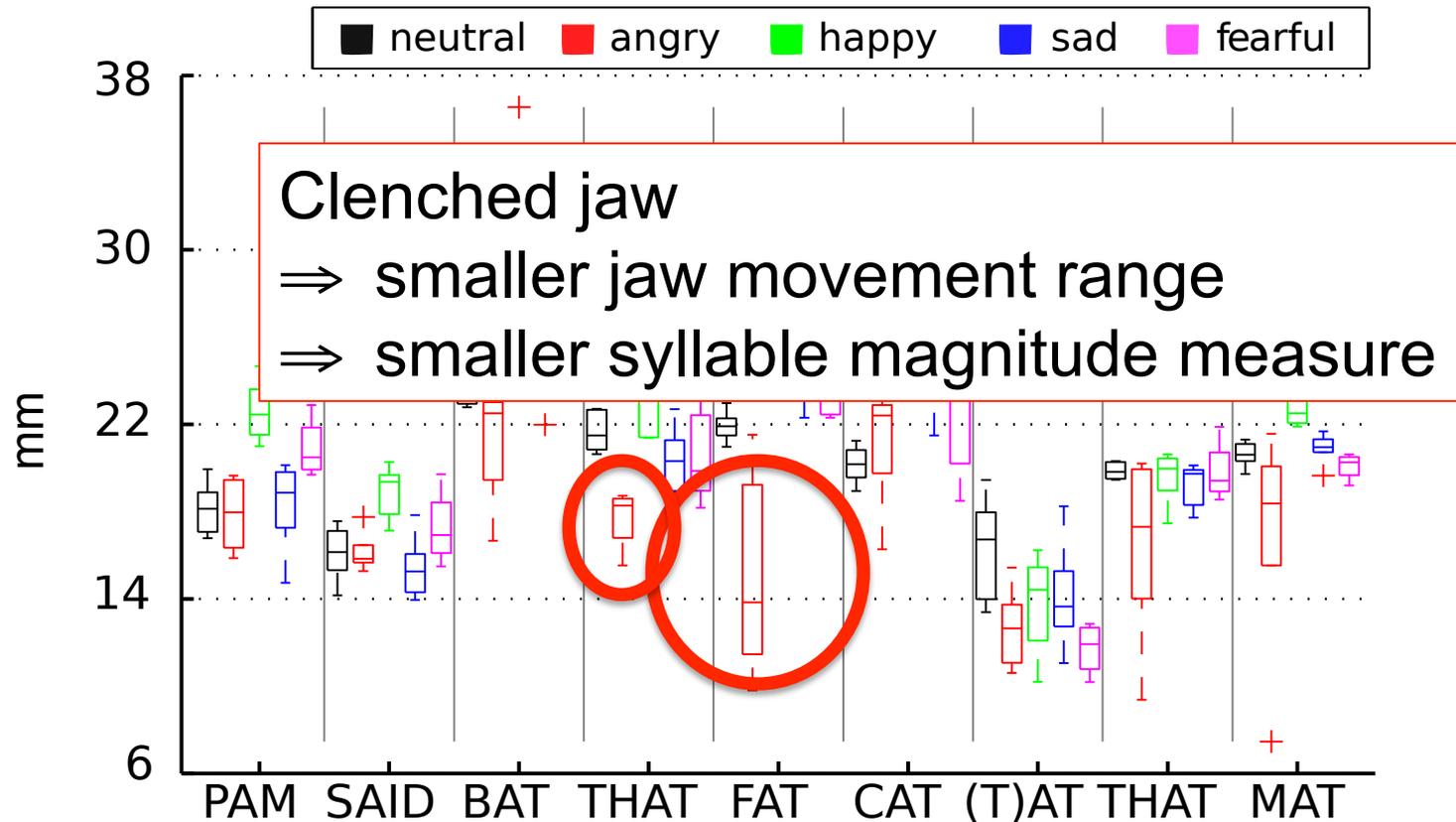
- Emotion affects syllable magnitude systematically



*Errorbar of syllable magnitude for each monosyllabic word and each emotion in the utterance*

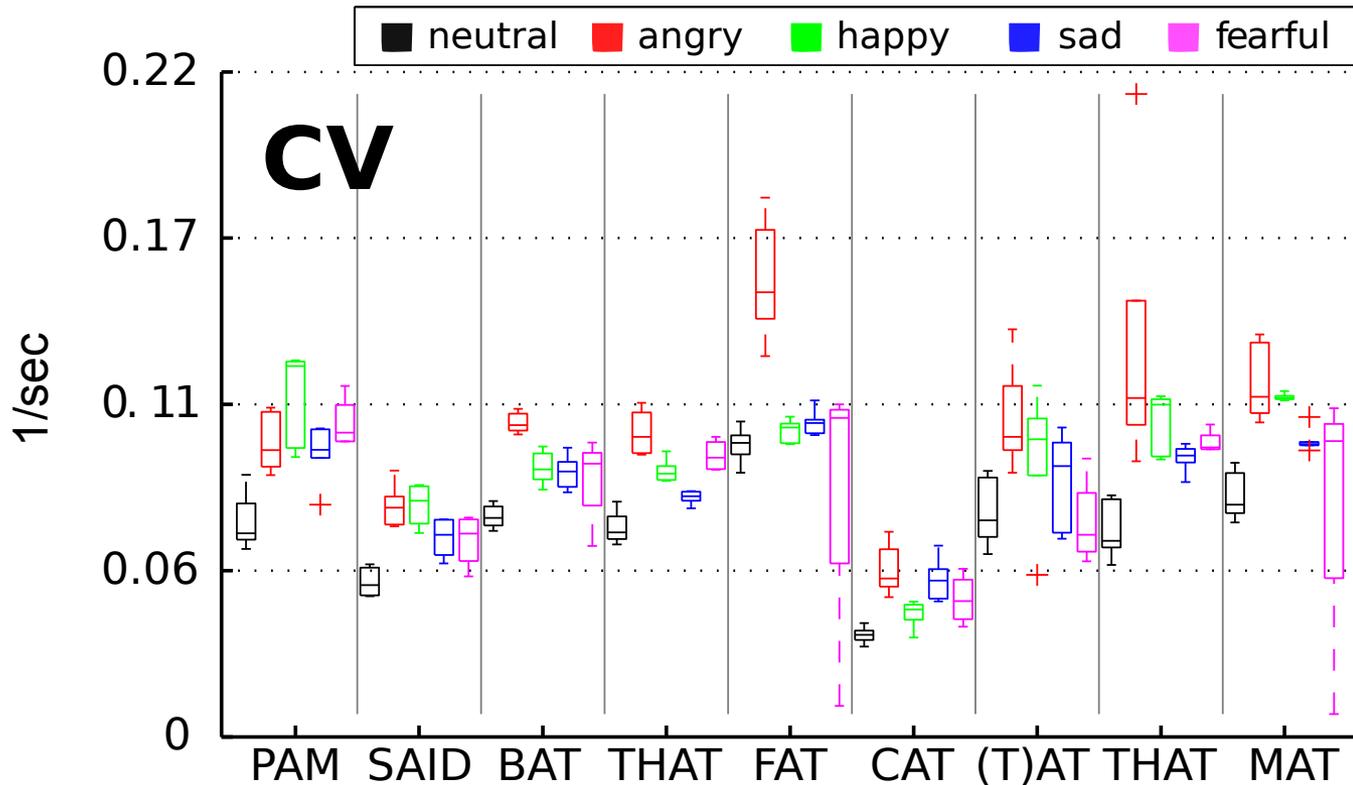


- Emotion affects syllable magnitude systematically



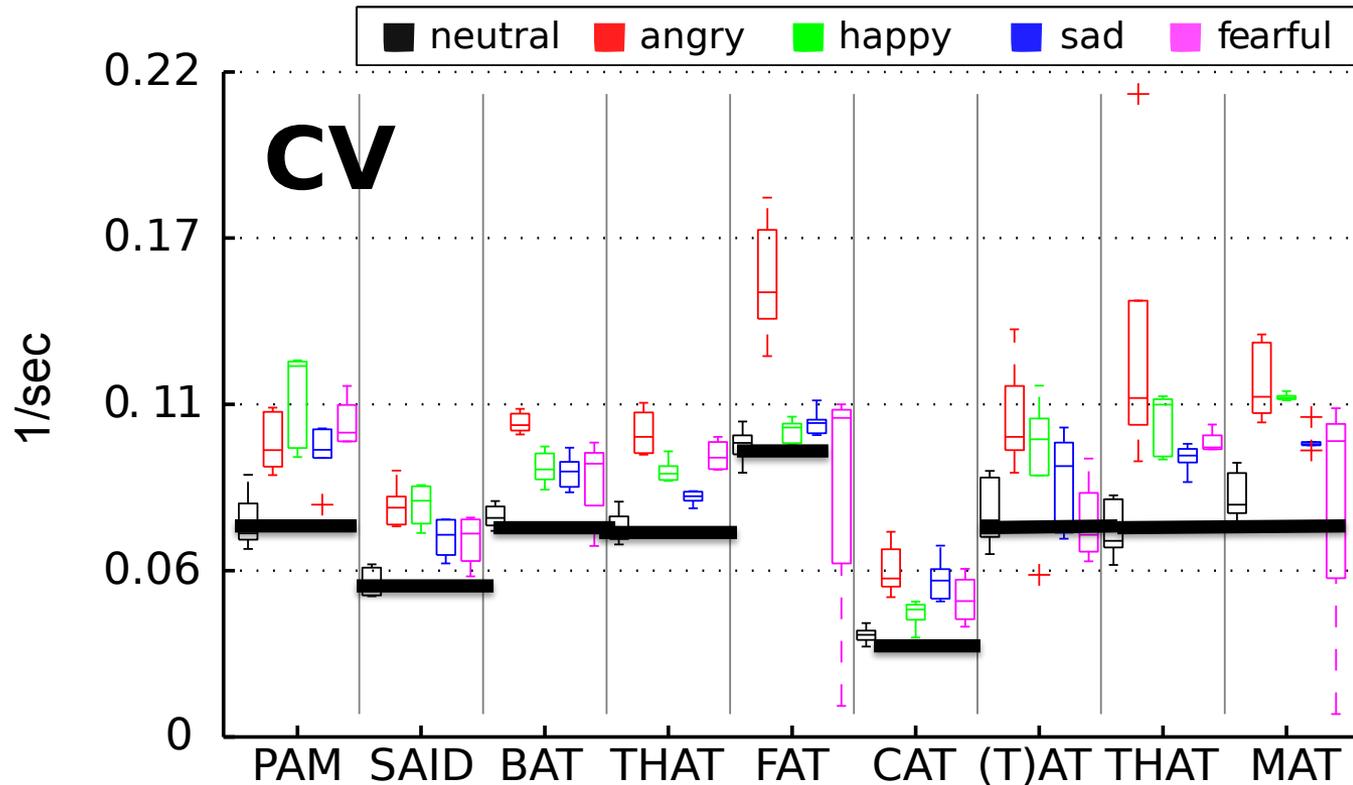
*Errorbar of syllable magnitude for each monosyllabic word and each emotion in the utterance*

- Emotion affects amplification (IRF amplification).



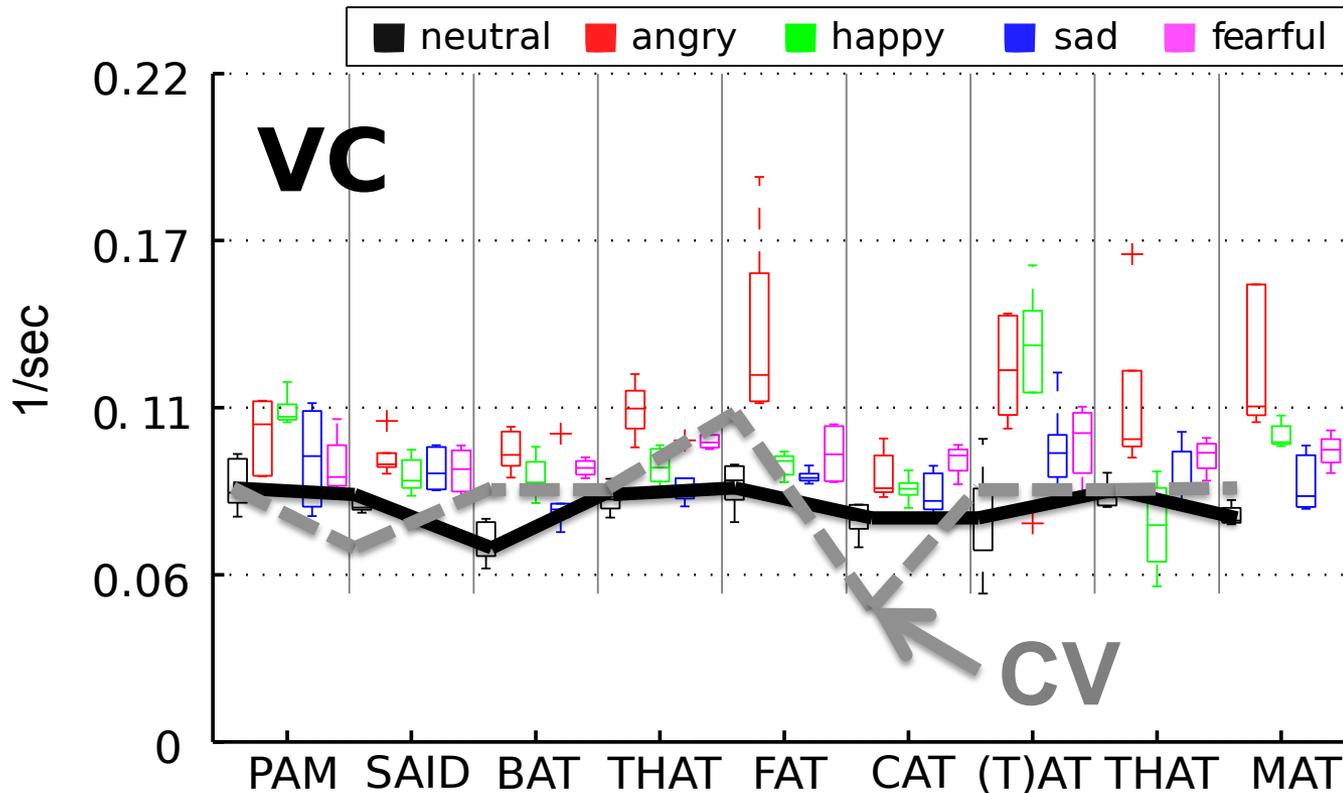
*Errorbar of the ratio of articulatory speed (at the iceberg point) to the syllable magnitude for each CV*

- Emotion affects amplification (IRF amplification).



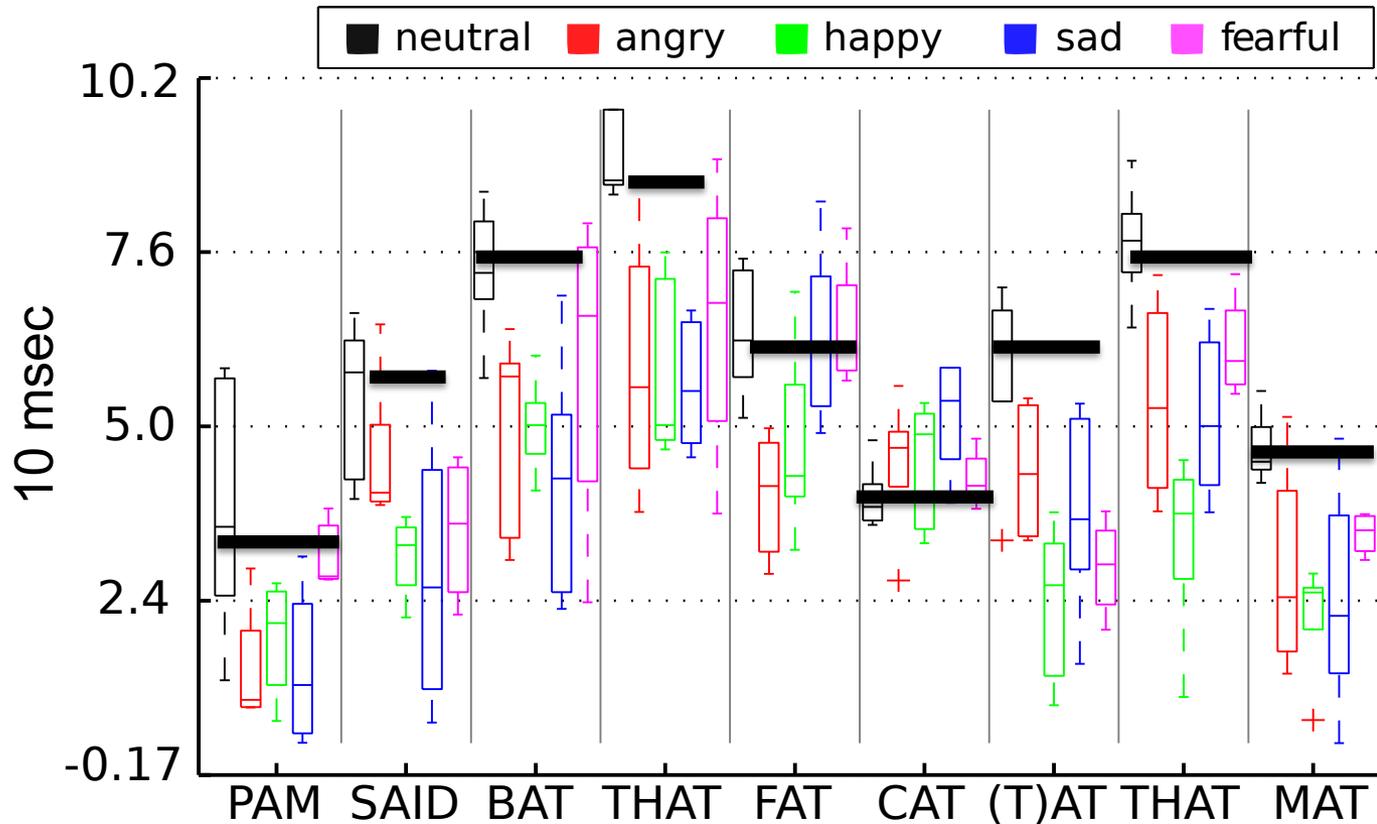
*Errorbar of the ratio of articulatory speed (at the iceberg point) to the syllable magnitude for each CV*

- Critical articulator's movements for VC is more dependent on the syllable magnitude than for CV.



*Errorbar of the ratio of articulatory speed (at the iceberg point) to the syllable magnitude for each VC*

- Emotion affects timing ( $\leq$ IRF timing).



*Errorbar of the time difference between the onset pulse point and the iceberg point for each syllable*

- Emotion does not affect the strong linear relationship between articulatory speed and movement range of critical articulator for consonant.  
=> Consistent stiffness across syllable
- Emotion affects shadow angle: The relation between the jaw excursion and the articulatory duration.
- Emotion also affects amplification and timing of the consonant gestural controls.
- Need more comprehensive analysis and improvement of the current C/D model for more realistic representation of emotional speech

- The analyzed parameters are not independent from each other. Hence, a joint analysis and modeling for the variation of these parameters as a function of emotion is important to understand emotional variability in the C/D model framework.
- Articulatory re-synthesis experiment with emotion conversion can be useful for evaluating the joint model.

- This work was supported by NSF IIS-1116076, NIH DC007124, and Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (C)#22520412 and (C)#25370444. Special thanks to Mary Francis for her devotion and help in all SAIL research efforts. Many thank to Dr. Osamu Fujimura, Dr. Patrizia Bonaventura and Dr. Caroline Menezes for sharing your insights.

