Detailed Study of Articulatory Kinematics of Critical Articulators and Non-critical Articulators of Emotional Speech

Jangwon Kim, Sungbok Lee, Shrikanth S. Narayanan
jangwon@usc.edu, sungbokl@usc.edu, shri@sipi.usc.edu

[Supported by NIH and NSF grants]

Introduction

Hypothesis: Critical articulators and non-critical articulators are weighted differently for achieving distinctive emotional goals on top of linguistic goals at the same time.

Speakers may use the variability of non-critical articulators more than critical articulators for achieving non-linguistic goals, like emotion encoding.

Analysis List:

(1) How much emotional information is contained in the movements of critical articulators and non-critical articulators? Which has more?

(2) How differently are the articulatory kinematics of critical articulators and non-critical articulators correlated with emotion?

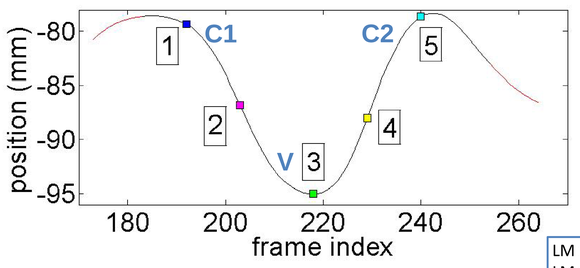
(3) What is the variability of critical articulators and non-critical articulators associated with distinctive emotion encoding?

Experimental setup

Electromagnetic Articulography database

- 5 acted emotions, 1 male (SB) + 2 females (JN, JR)
- Unit of segments: syllable (CVC)
  - Same critical articulator for the two consonants.
    Ex) [p iy p], [f ay v], [t ay t], [n ay t]
  - 38 parameters: position, speed, acceleration at 5 landmark (LM) points, statistics of articulatory movements in LM-based regions

5 landmark points



- Number of instances for analysis

		JN	JR	SB
Critical articulator	Tongue tip	422	498	389
	Lips	250	322	253

LM 1: releasing start
LM 2: maximum speed from LM 1 ~ LM 3
LM 3: largest opening
LM 4: maximum speed from LM 3 ~ LM 5
LM 5: closure point
LM 1 ~ 3: releasing
LM 3 ~ 5: approaching

Emotion classification for analysis (1)

- Brute-force feature selection and classification (SVM)
- Features: all 38 parameters (by chance: 20%)

	Tongue tip			Lips		
	JN	JR	SB	JN	JR	SB
CA	41.9	44	54.8	52.8	56.8	62.1
NCA	49.2	46.6	57.3	63	53.8	57.1

CA: critical articulator, NCA: non-critical articulator

- More distinctive power on lip movement parameters
- For tongue tip, more emotional information when NCA

- Confusion matrix (SB for example)

		Classified									
		Tongue tip					Lips				
		Neu	Hot A	Cold A	Hap	Sad	Neu	Hot A	Cold A	Hap	Sad
Evaluated	Neu	22	12	7	19	7	19	8	10	3	9
	Hot A	17	33	5	9	15	6	25	3	9	2
	Cold A	0	0	78	7	3	5	5	33	3	12
	Hap	8	7	8	49	13	1	16	6	17	4
	Sad	5	19	3	15	31	2	0	3	1	51

Neu: neutrality, Hot A: hot anger, Cold A: cold anger, Hap: happiness, Sad: sadness

ANOVA for analysis (2)

Statistics on LM-based regions

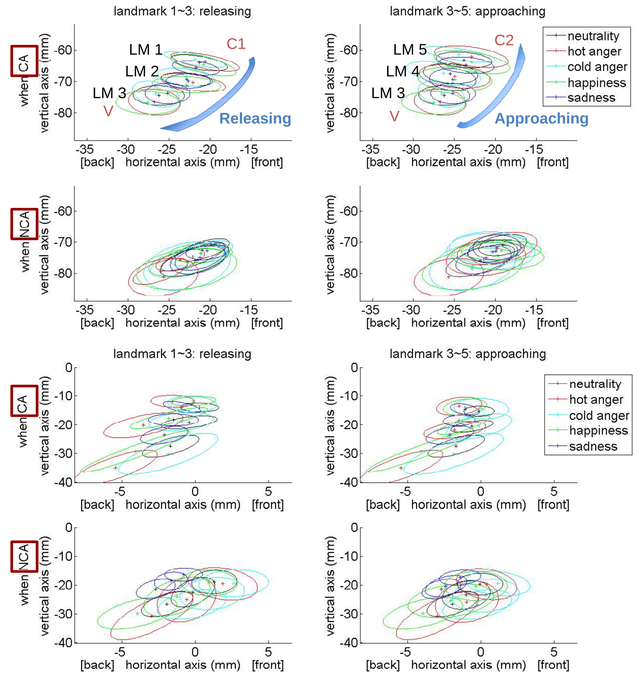
- Overall characteristics of articulatory kinematics of articulatory trajectory of critical articulators are emotionally significant.
  - Mostly significant for five categorical emotions
  - However, average positions of critical articulators in the horizontal and vertical axes tend to be less significant or not significant ($p > 0.05$).
- Significance of non-critical articulators tend to be higher than that of critical articulators.

Position (x, z), speed and acceleration at each LM

- Vertical positions at LM 4 and LM 5 (approaching region) tend to be less significant.
- Tangential speed and acceleration are mostly significant at LM 2 and LM 4.
- Almost all lip parameters are significant. They tend to be more significant than those of tongue tip's in general.
- Position parameters are most significant at the largest opening point (LM 3).
- Tangential speed and acceleration are most significant at the maximum speed points in releasing and approaching regions.

Distribution plots for analysis (3)

- Sample distribution plots (2σ ellipsoid) for emotions
- Releasing (LM 1 ~ LM 3) and approaching (LM 3 ~ LM 5)



- More variation of positioning of NCA, especially in the horizontal axis even at LM 1 and LM 5 (closure points).
- The emotional variation of CA tends to be larger as reaching the largest opening point. The emotional variation of NCA is more universal in the whole syllable region.
- In general, variation of articulatory positions at the 5 landmarks follows the property of emotion in the arousal dimension.
- TT: more variability on approaching movement (especially, horizontal direction) than releasing movement.

Conclusions & future work

- Distinctive emotional information: NCA > CA for TT. (? for lips)
- More universal emotional encoding patterns of NCA than CA
- Variation of articulatory positions in a syllable (CVC) region follows the property of emotion in the arousal dimension
- Future works:
  - Articulatory target for linguistic goals: region or contour?
  - Inter-speaker comparison, phase, etc.