USC Viterbi
School of Engineering

Ming Hsieh Department of Electrical Engineering

A study of emotional information present in articulatory movements estimated using acoustic-to-articulatory inversion

Jangwon Kim, Dept. of Electrical Engineering, Signal Analysis and Interpretation Lab.

Motivation & Introduction

Motivation: Difficulty of articulatory data collection of emotional speech, the (potential) benefits of the estimated articulatory data (complementary information gain to other emotional features from acoustic signal)

Goal: Understanding the effectiveness of acoustic-to-articulatory inversion technique for emotional speech.

Research Questions (RQ):
(1) How similar is the estimated trajectory to the original articulatory trajectory for various emotions?
(2) How much does the estimated trajectory preserve emotion specific information (ESI)?

Experimental Setup

For RQ1: comparison of two articulatory trajectories (original-estimated) using the MSE measure of the same emotion pair with that of different emotion pair.
For RQ2: to test using emotion classification accuracy.

Generalized smoothness criterion based inversion
-Acoustic-to-articulatory inversion used here
-Constraining individual articulatory trajectories by the corresponding articulator-specific smoothness factors.

Database: Electromagnetic articulography
-Parallel waveform and articulatory trajectory
-5 categorical & elicited emotions
-2 females, 1 male spoke 7 sent, 3 styles, 4~5 rep.
-6 articulatory points (3 for tongue, 2 for lips, 1 for jaw)

Conclusions & Future Work

Conclusions:
-Articulatory data estimated using GSC based inversion carry important ESI, but smaller than original.
-ESI in the estimated trajectories offers complementary emotion information to that in the acoustic prosodic features.
-Remaining Question: how ESI is encoded in the estimated articulatory movements (in dimensional or categorical way of emotion?)

Future work directions:
-Investigate other inversion techniques
-Subject-independent inversion for emotional speech

## Results and Discussion

Average MSE of two trajectories:
- Results of a female as an example

			Original				
			Emotion				
			Neu	HAng	CAng	Hap	Sad
Original	Emo-tion	Neu	2.26	3.82	2.81	3.52	2.91
		HAng	3.82	3.62	3.69	3.88	3.85
		Cang	2.81	3.69	2.72	3.55	2.95
		Hap	3.52	3.88	3.55	3.14	3.50
		Sad	2.91	3.85	2.95	3.50	2.90
	aveDiff		-1.01	-0.19	-0.53	-.0.47	-0.40
	F		335	10	218	160	80
	P-value		0.000	0.001	0.000	0.000	0.000

			Original					ave-Diff	F	P-value
			Emotion							
			Neu	HAng	CAng	Hap	Sad			
Estimated	Emo-tion	Neu	2.46	3.99	2.93	3.36	2.66	-0.78	789	0.000
		HAng	3.00	3.77	3.27	3.41	3.01	0.60	824	0.000
		Cang	2.64	3.92	2.94	3.36	2.68	-0.21	39	0.000
		Hap	3.04	4.02	3.36	3.41	3.01	0.05	7	0.008
		Sad	2.86	4.04	3.12	3.41	3.01	-0.64	493	0.000
	aveDiff		-0.43	-0.22	-0.24	0.03	-0.12			
	F		306	67	136	4	16			
	P-value		0.000	0.000	0.000	0.046	0.000			

-Emotional contrast in terms of average MSE for esti-orig articulatory pare case (bottom table) is less than that for the two orig articulatory pair case (top table).
-Inversion does not well maintain the nature of emotional contrast of orig articulatory trajectory in the arousal dimension.

Emotion Classification:
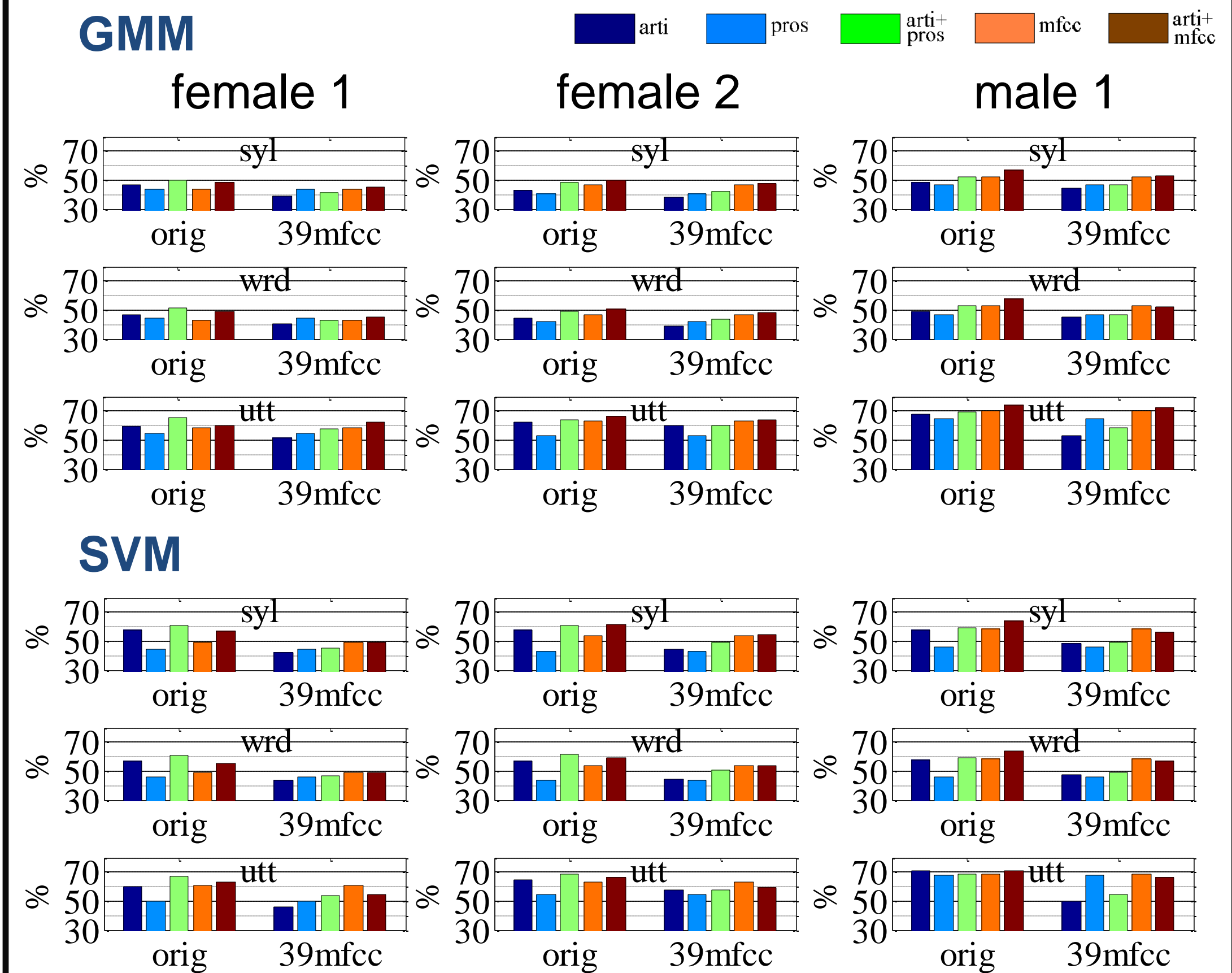
-Estimated articulatory data has emotion-specific information (ESI) but not to the same degree as in orig. articulatory data.
-Estimated articulatory data has ESI complementary to the acoustic and prosodic features.