INTERPLAY BETWEEN LINGUISTIC AND AFFECTIVE GOALS IN FACIAL EXPRESSION DURING EMOTIONAL UTTERANCES

Carlos Busso and Shrikanth S. Narayanan

Speech Analysis and Interpretation Laboratory (SAIL)
Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089
busso@usc.edu, shri@sipi.usc.edu

Abstract

Communicative goals are simultaneously expressed through gestures and speech to convey messages enriched with valuable verbal and non-verbal clues. This paper analyzes and quantifies how linguistic and affective goals are reflected in facial expressions. Using a database recorded from an actress with markers attached to her face, the facial features during emotional speech were compared with the ones expressed during neutral speech. The results show that the facial activeness is mainly driven by articulatory processes. However, clear spatial-temporal patterns are observed during emotional speech, which indicate that emotional goals enhance and modulate facial expressions. The results also show that the upper face region has more degrees of freedom to convey non-verbal information than the lower face region, which is highly constrained by the underlying articulatory processes. These results are important toward understanding how humans communicate and interact.

Introduction

Motivation

- Linguistic and affective goals modulate speech and gestures to convey the desired messages
- Linguistic and affective goals co-occur during human interaction, sharing the same channels
- In spite of all this emotional modulation, the linguistic goals are successfully fulfilled
- Some internal control needs to buffer, prioritize and execute these communicative goals

Hypotheses

- Linguistic and affective goals interplay as primary and secondary controls
- During speech, linguistic goals are prioritized over affective goals
- As results, emotional "fingerprint" in facial expressions is not steady distributed
- Some facial areas have more degree of freedom to display non-verbal clues

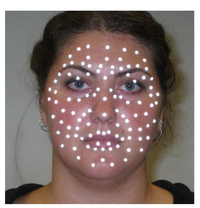
Purpose of the analysis

- To investigate the interplay between linguistic and affective goals in facial expressions
- To quantify the degree of freedom of facial areas to express the emotion
- To quantify articulatory constraints imposed in the face during active speech

Methodology

Audio-visual database

- Database recorded from an actress with markers attached to her face (404 sentences)
- She read a corpus four times expressing sadness, happiness, anger and neutral state
- Markers were tracked with a VICON motion capture system with three cameras
- Position of the facial markers were corrected by compensating head translation and rotation
- The audio was simultaneously recorded with a SHURE microphone

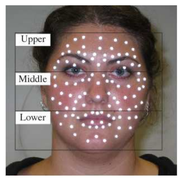


(a) Facial marker layout

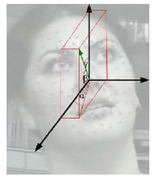
(b) Motion capture system

Feature extraction

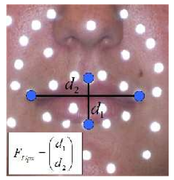
- Face parameterization
  - Head motion directly derived from the estimated rotational matrix
  - Eyebrow motion computed by subtracting the position of two markers chosen in right eye
  - Lip motion estimated with a 2-D feature vector, which measures the opening of the mouth
- In addition, each facial marker is considered as a facial feature
  - Results of the markers are aggregated in three areas: upper, middle and lower face regions



(a) Facial marker division



(d) Head motion features



(e) Lip features

Emotional modulation

Temporal emotion modulation

- Mean and variance of utterance durations for emotional classes were higher than for neutral
- The speaking rate had higher average values during emotional speech
- The vowels durations mean for anger and happiness were higher than for sadness and neutral

Spatial emotional modulation

Facial activation analysis

- During active speech, areas in the face present different levels of movement activity
- The *motion coefficient* Ψ is used to measure facial activeness

$$\Psi_u = \frac{1}{T_u} \sum_{i=1}^{T_u} D_{eq}(\vec{X}_i^u, \vec{\mu}^u) \quad (1)$$

- The lower face region has the highest activeness levels
  - Articulatory processes play a crucial role in face expressions
  - Linguistic goals have priority during active speech
- Emotional modulation affects the activeness in the face
  - Activeness for happiness and anger is more than 30% higher than in the neutral case
  - The activeness in the lower face region for sadness decreases 20% compared with neutral
- Activeness in upper face region increases more then other regions
  - Valuable non-verbal information is conveyed in this area

Facial Area	N	S	H	A
Head Motion	1.92	4.11	4.65	4.53
Eyebrow	0.05	0.06	0.11	0.12
Lip	4.51	3.55	6.22	7.28
Upper region	0.66	0.79	1.47	1.47
Middle region	0.88	0.88	1.43	1.56
Lower region	3.15	2.46	4.21	4.59

(a) Neutral

(b) Sad

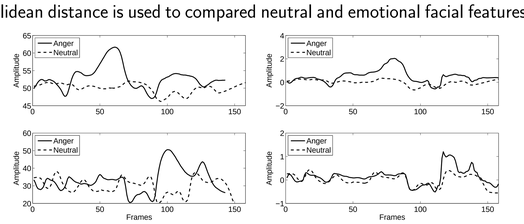
(c) Happy

(d) Angry

Neutral vs. emotional analysis

- Compare facial expressions displayed during neutral and emotional utterances with similar semantic content (linguistic content is fixed)
- Dynamic time warping (DTW) is used to align the sentences
- Correlation and Euclidean distance is used to compared neutral and emotional facial features

(a) Alignment path

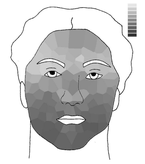


(b) Original lip features

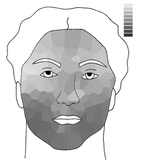
(c) Warped lip features

Correlation analysis

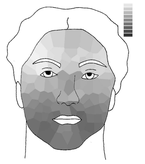
- Correlation was calculated between the neutral and warped emotional facial features
- Higher correlation implies higher articulatory constraints (follow the speech)
- Lower facial region presents the highest correlation levels
  - More constrained by underlying articulation
- Upper facial region has the lower correlation levels
  - it can communicate non-verbal information regardless of the linguistic content



(a) Neutral-sad



(b) Neutral-happy



(c) Neutral-angry

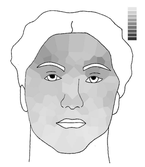
Facial Area	Pearson's correlation			Euclidean distance		
	Neu-Sad	Neu-Hap	Neu-Ang	Neu-Sad	Neu-Hap	Neu-Ang
Head Motion	0.24	0.25	0.17	4.28	3.83	3.44
Eyebrow	0.25	0.15	0.07	0.69	2.56	1.31
Lip	0.54	0.50	0.53	0.38	1.61	0.82
Upper region	0.27	0.24	0.15	1.08	2.49	2.02
Middle region	0.46	0.38	0.37	0.63	2.12	1.27
Lower region	0.58	0.52	0.53	0.46	0.95	0.71

Euclidean distance analysis

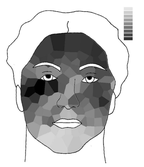
- Facial features were normalized before estimating Euclidean distance

$$\widehat{F}_t^{(neu)} = (F_t^{(neu)} - \vec{\mu}_t^{(neu)}) \cdot \alpha_t^{(neu)} \quad (2)$$
$$\widehat{F}_t^{(emo)} = (\widetilde{F}_t^{(emo)} - \vec{\mu}_t^{(neu)}) \cdot \alpha_t^{(neu)} \quad (3)$$

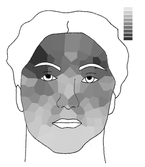
- High values indicate that facial features are more independent of the articulation
- Upper face region presents the highest differences between the neutral and emotional facial expression
  - Emotional goals control this area



(a) Neutral-sad



(b) Neutral-happy



(c) Neutral-angry

Discussions and conclusions

- During speech, facial activeness is mainly driven by articulation
- The activeness levels are affected by emotional modulation
  - Linguistic and affective goals co-occur during active speech
- There is interplay between linguistic and affective goals in facial expression
  - Upper/middle face region have more degree of freedom to convey non-verbal messages
  - The lower face region is more constrained by articulation
- The results have important implications in areas such as emotion perception, emotion recognition, speech production and facial animation
  - Upper face region is perceptively the most important facial region to detect visual prominences [Swerts, 2006] [Lansing, 1999]
  - Upper and lower face regions are sufficient to accurately recognize emotions [Busso, 2004]
  - These facial areas should be properly modeled to convey more realistic animations

Future work

- Analyze what happens when conflict between these communicative channels occur
  - Other gestures with more degree of freedom may be used (e.g. eyebrow, head motion)
- Analyze facial expressions during acoustic silence
  - Affective goals may have priority
- Analyze how to model this spatial-temporal emotional modulation
- Generalize and validate these results with a database recorded from more subjects