AFFECT RECOGNITION FROM LIP ARTICULATIONS

Rizwan Sadiq and Engin Erzin
Multimedia Vision and Graphics Lab, Koç University Istanbul, Turkey

Introduction

Main Questions:

1. How significant are lip articulations for affect recognition?

2. Does all phonetic classes play equal role in affect recognition from lip articulations?

Our Solution:

- Find phonemes that can better discriminate affect from lip articulations
- Test affect recognition with and without selected phonemes

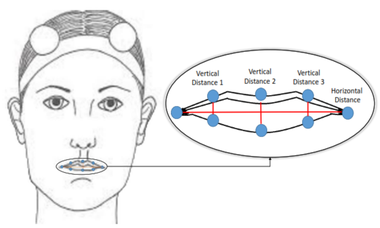
Dataset

- IEMOCAP: Interactive emotional dyadic motion capture database
- Affect attributes: Activation, Valence, Dominance
- Three Discrete Affect Levels: A1, A2 and A3

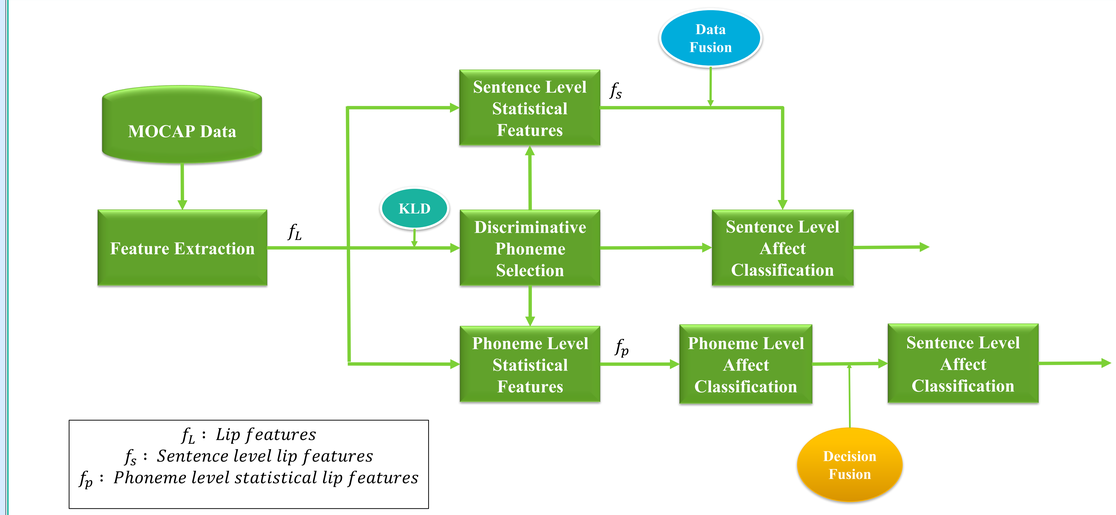
$$\begin{Bmatrix} A1 & if\ 1 \le A \le 2\text{: } Low\ \ AVD \\ A2 & if\ 2 < A < 4\text{: } Medium\ AVD \\ A3 & if\ 4 \le A\ \ \ \text{: } High\ AVD \end{Bmatrix}$$

Feature Extraction

- f_L : One horizontal and three vertical lip distances, extracted at frame level to create 4-dimensional feature vector ()
- f_S & f_P: 11 Statistical functionals over temporal windows (sentence & phoneme) to define 44-dimensional segment level features



System Overview



Discriminative Phoneme Selection

- Define a symmetric KullbackLeibler divergence (KLD) of lip features $\boldsymbol{f}$ given affect state $\boldsymbol{A}$ and phoneme $\boldsymbol{p}$:

$$D_{mn}(f|p) = KLD(P(f|A_m, p), P(f|A_n, p))$$

- For each phoneme, compute cumulative distance between affect classes:

$$S(f|p) = w_p(D_{12}(f|p) + D_{23}(f|p) + D_{13}(f|p))$$

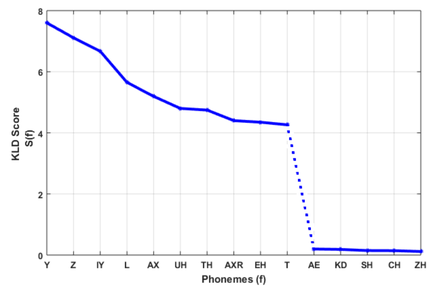
- Higher the distance function $S(f|p)$, better the discrimination

Affect Classification

- 3-Level discrete affect classification
  - Using SVM from LIBSVM package
- Affect classification was performed on sentence level using two approaches
  - Data fusion
    - Constructed segment level lip feature vectors over all duration of the sentence
  - Decision Fusion
    - Constructed classifier for each phoneme
    - Used majority voting over the phoneme level classifier outputs

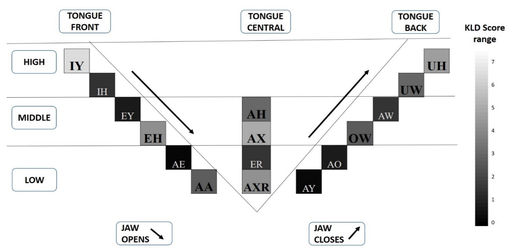
Results on Discriminative Phonemes

Cumulative KLD score of top 10 and lowest 5 phonemes:



Analysis over Vowels

Among the vowels, the least discriminative region is observed as jaw opens and tongue is at back:



Analysis over Consonants

- Top two discriminative consonants:
  - The voiced palatal **Y**
  - The voiced alveolar **Z**
- The least discriminative consonants:
  - The labio-dentals **F** and **V**
  - Alveo-palatals **SH**, **ZH, CH, JH**

	Bilabial	Labio-Dental	Inter-Dental	Alveolar	Alveo-Palatal	Palatal	Velar	Glottal
Stop (oral)								
voiceless	P			T			K	
voiced	B			D			G	
Nasal (stop)	M			N			NG	
Flaps				DD	TD			
Fricative								
voiceless		F	TH	S	SH			HH
voiced		V	DH	Z	ZH		IX	
Affricate								
voiceless				TS	CH			
voiced					JH			
Glide								
voiceless	M						M	HH
voiced	W					Y	W	
Liquid				DX	L	R		

KLD Score range

AVD Classification Results

Comparison of weighted (WA) and unweighted (UA) classification accuracies of AVD attributes with selected phonemes and all phonemes using data fusion:

	Classification Accuracy (%)					
	Selected Phoneme			All Phonemes		
	%Val	%Act	%Dom	%Val	%Act	%Dom
WA	**46.44**	72.13	64.47	45.43	71.40	61.79
UW	**42.33**	41.42	**39.61**	38.82	40.37	38.53

Comparison of weighted (WA) and unweighted (UA) classification accuracies of AVD attributes with selected phonemes and all phonemes using decision fusion:

	Classification Accuracy (%)					
	Selected Phonemes			All Phonemes		
	%Val	%Act	%Dom	%Val	%Act	%Dom
WA	46.16	**72.16**	**64.92**	46	72	64.7
UW	36.33	**42.65**	38.98	35	40	38.1

Conclusions

- Lip features can attain affect recognition performance above the chance level
- Selected list of discriminative phoneme articulations can better classify all affect attributes using only the lip features

Future Work

- Lip articulations can be used in multimodal affect recognition systems
- Affective lip syncing

Email: rsadiq13@ku.edu.tr