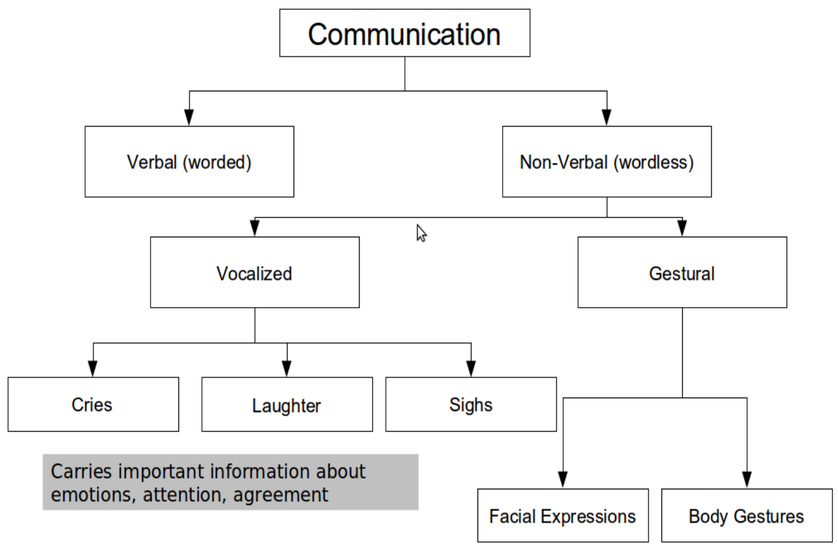
Ming Hsieh Department of Electrical Engineering

Classification of emotional content of sighs in dyadic human interactions

Rahul Gupta, Chi-Chun Lee, Shrikanth Narayanan
Signal Analysis and Interpretation Lab

USC Viterbi School of Engineering

Human Communication



Motivation and Hypotheses

Sighs: Audible deep intake and release of breath

Motivation

- Understanding the role of Non-Verbal Vocalizations in human communication
- Investigating the role of sighs in particular in emotional expression

Hypotheses

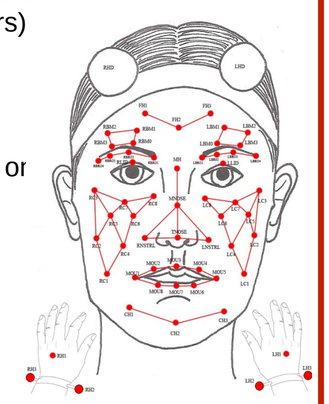
- Sighs occur with respect to the emotional state of the person [1,2,3]
- Sighs can be classified into positive and negative emotions based on their acoustic and gestural properties [1]

[1] D. Sauter, , F.Eisner, , A. J.Calder, , and S. K. Scott, "Perceptual cues in nonverbal vocal expressions of emotion"Quarterly Journal of Experimental Psychology, 63(11), 2251-2272, 2010

Database

IEMOCAP Database

- 10 actors, 5 sessions, dyadic spoken interaction (~12 hrs)
- Vicon Motion Captured of Facial & Hand Movement
  - 61 Markers on one subject (55 on face, 3 on each hand)
  - (x,y,z) coordinate captured at 120Hz
- Audio, transcripts, forced alignment available
- Spontaneous dialog & Scripted plays
  - 8 hypothetical scenarios, 3 scripts



Annotation

Emotion label per utterance

- 3 evaluators
- Categorical (sad, happy, neutral etc..)
- Dimensional (valence, activation, dominance)

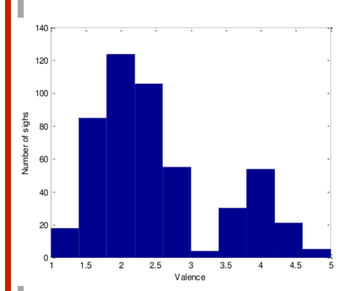
End and start time for sighs

- Evaluation done by single person

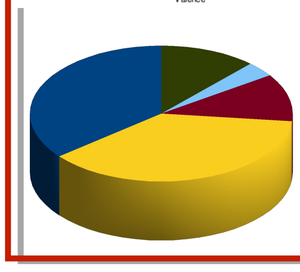
Valence for sigh approximated by the closest utterance by the same person

Experiment 1

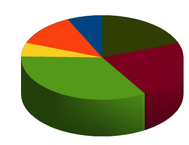
Hypothesis : Sighs occur with respect to a person's emotional state



Bimodal distribution of sigh valence used to classify sighs as belonging to +ve or -ve class

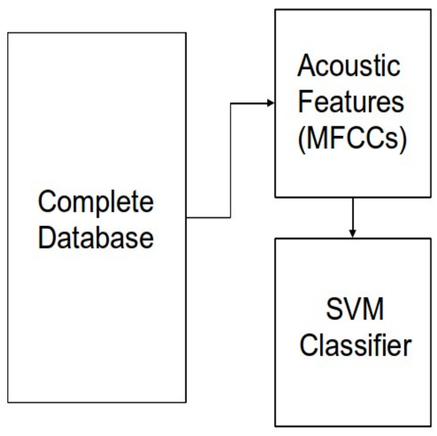


Negative Sighs
Total Number: 388
Positive Sighs
Total Number: 114



Experiment 2a

Using acoustic cues only

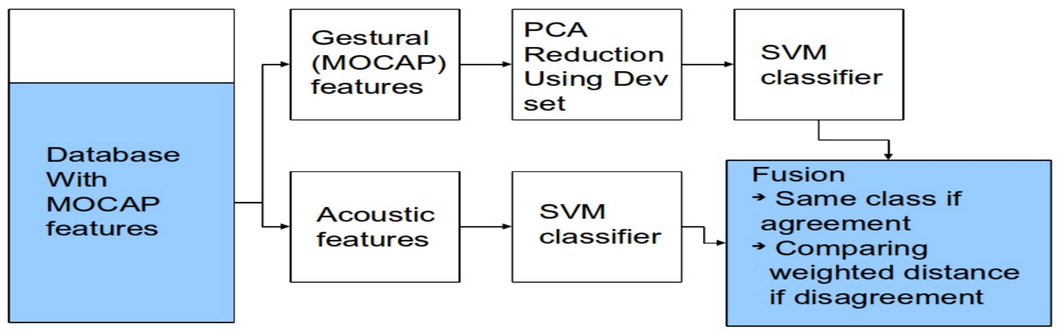


Model	Unweighted Accuracy
Chance	50%
SVM	57.51%

Experiment	Positive Class Accuracy	Negative Class Accuracy
Using MFCC features on the entire database	28.95	86.08

Experiment 2b

Using acoustic and gestural cues



$$\alpha(w_{acoustic}.x_{acoustic} - b_{acoustic}) > (w_{mocap}.x_{mocap} - b_{mocap})$$

Model	Unweighted Accuracy (%)
Chance	50%
SVM using acoustic models	55.74%
SVM using MOCAP features	57.66%
Fusion	58.26%

Experiment	Positive Class Accuracy	Negative Class Accuracy
SVM using acoustic features	24.19	87.30
SVM using MOCAP features	40.32	75.00
Fusion	40.32	76.19

Conclusions and Future Work

- Sighs are observed to occur with different emotions
- The features carry discriminative power between the positive and the negative class emotions

Future Work

- Applications to bigger databases for better understanding
- Analysis of other form of Non verbal vocalizations
- Incorporate in emotion recognition

References

[1] D. Sauter, F. Eisner, P. Ekman, and S. K. Scott, "Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations" in Proceedings of the National Academy of Sciences, 107(6), 2408-2412, 2010

[2] D. Sauter, , F.Eisner, , A. J.Calder, , and S. K. Scott, "Perceptual cues in nonverbal vocal expressions of emotion"Quarterly Journal of Experimental Psychology, 63(11), 2251-2272, 2010

[3] V. Reddy, E. Williams and A.Vaughan, "Sharing humour and laughter in autism and Down's syndrome" British Journal of Psychology, 93: 219–242

[4] J. L. Driver and J. M. Gottman, "Daily marital interactions and positive affect during marital conflict among newlywed couples" Family Process, 43: 301–314.

Ming Hsieh Institute
Ming Hsieh Department of Electrical Engineering