Víctor Ponce-López[1,2], Xavier Baró[1,2] , Sergio Escalera[2]

[1]Internet Interdisciplinary Institute, UOC, Carrer Roc Boronat 117, 08018, Barcelona.
[2]Dept. Matemàtica Aplicada i Anàlisi, UB, Gran Via de les Corts Catalanes 585, 08007, Barcelona
vponcel@uoc.edu, xbaro@uoc.edu, sergio@maia.ub.es

Multi-modal Non-Verbal Communication Analysis in Victim-Offender Mediation Sessions

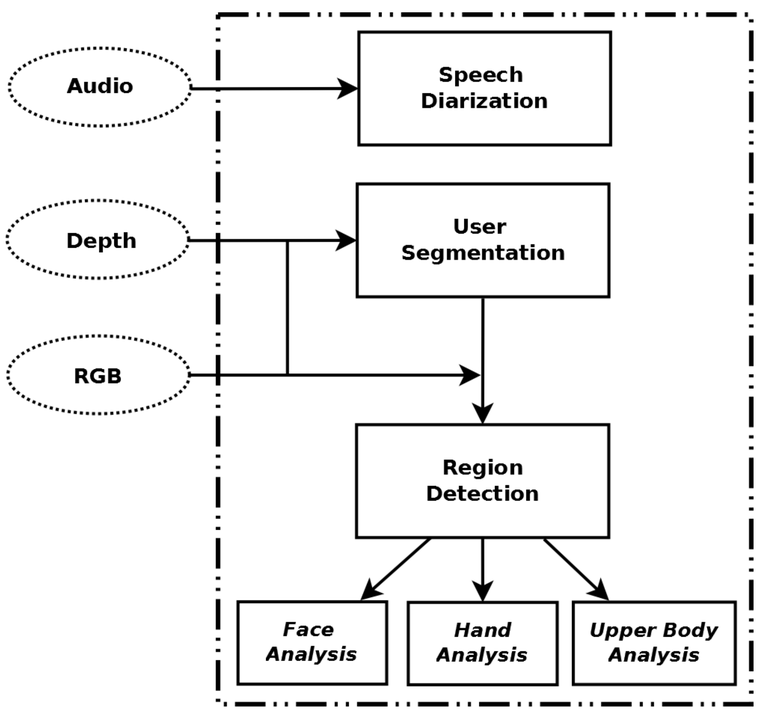
ABSTRACT

Multi-modal feature extraction using computer vision is applied to identify objective behavioral indicators on the different parts involved in Victim-Offender Mediations. The receptivity, agreement, and satisfaction reached among these parts is predicted using social signal processing and machine learning approaches according to the expert opinions.

1. Motivation

- Human language is essential in human social interactions.
- Non-verbal communication is found within the human language through the gestures, and beyond the human speech [1,5].
- Multi-modal technologies allow to capture audio-RGB-depth data from conversational scenarios to analyze behavioral indicators appearing on the subjects [3].

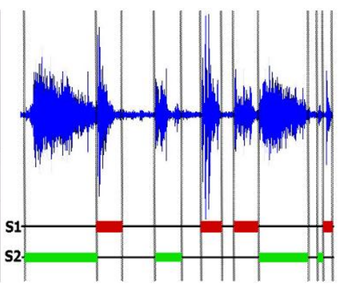
2. System



Face detection and Head Pose Estimation [4]

Depth Map [6]

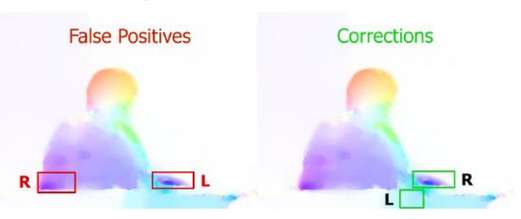
Speech signal and diarization [3]



$$P(l|I,\dot{p}) = \frac{1}{T}\sum_{t=1}^{T} P_t(l|I,\dot{p}) \quad \textbf{[6]}$$

Heuristic correction examples

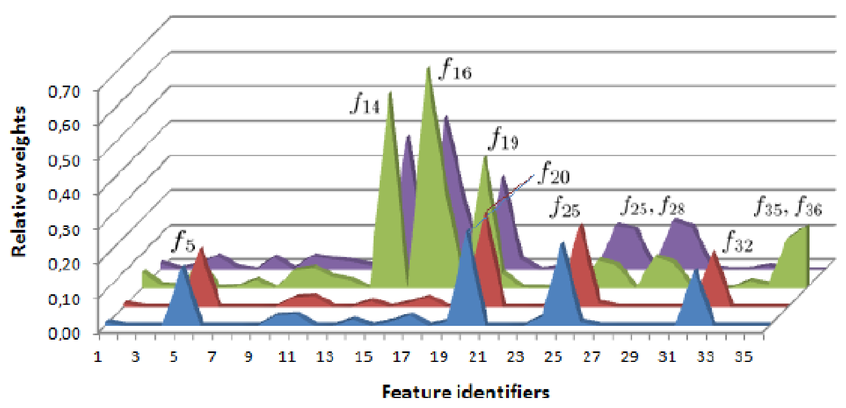
Correction of facial fitting



Correction of hand positions

3. Results

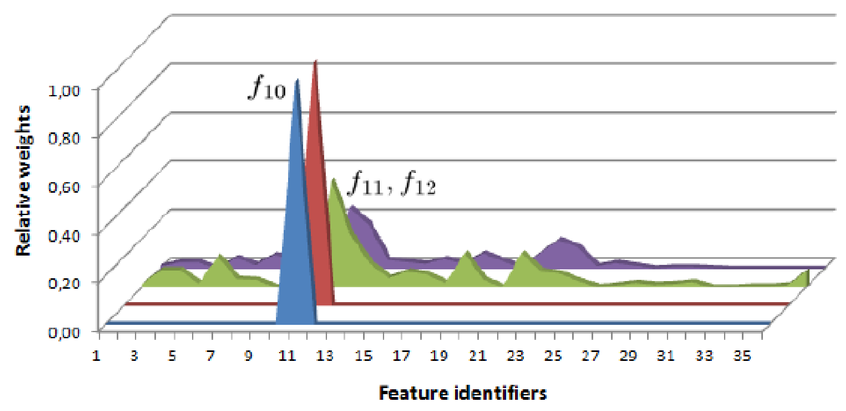
Weighted feature selection predicting the Receptivity



Most relevant features

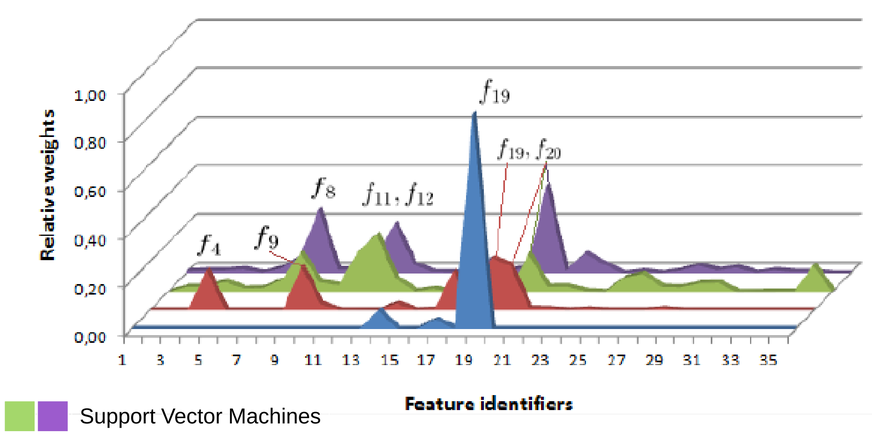
Agitation of upper body and hands while looking

Weighted feature selection predicting the Agreement



Age

Weighted feature selection predicting the Satisfaction



Hands agitation while looking

Support Vector Machines

Adaboost

Prediction Accuracy using statistical machine learning approaches

Label	Adaboost	LDA	PNN	CF	FF	SVM
Receptivity	**82%**	50%	39%	71%	79%	**82%**
Agreement	71%	32%	29%	71%	**75%**	71%
Satisfaction	**79%**	54%	21%	75%	**79%**	**79%**

4. Conclusion

- Proposed a multi-modal framework for the analysis of non-verbal communication in Victim-Offender Mediations.
- Usability of computer vision, signal processing, and machine learning strategies in conversational processes is shown with results upon 82% of accuracy.

References

[1] A. Pentland. *Honest Signals: How They Shape Our World*. The MIT Press, Massachusetts, 2008.
[2] A. Marcos-Ramiro, D. Pizarro-Perez, M. Marron-Romera, L. Nguyen, and D. Gatica-Perez. *Body comunicative cue extraction for conversational analysis*. FG, 2013.
[3] S. Escalera, O. Pujol, P. Radeva, J. Vitrià, and Maria T. Anguera. *Automatic detection of dominance and expected interest*. EURASIP Advances in Signal Processing, Research Article, 2010.
[4] X. Zhu and D. Ramanan. *Face detection, pose estimation and landmark localization in the wild*. CVPR, 2012.
[5] D. McNeill. *Gesture and Thought*. University of Chicago Press, Chicago, 2005.
[6] J. Shotton, A.-W. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. *Real-time human pose recognition in parts from single depth images*. CVPR, pages 1297-1304, 2011.