# SPEECH EMOTION RECOGNITION USING AMPLITUDE MODULATION PARAMETERS

*Arianna Mencattini[1], Eugenio Martinelli[1], Giovanni Costantini[1], Massimiliano Todisco[1], Barbara Basile[2], Marco Bozzali[2], Corrado Di Natale[1]*

[1]Dept. of Electronic Engineering, University of Rome Tor Vergata, Via del Politecnico 1, 00133 Rome, Italy, e-mail costantini@uniroma2.it
[2]Neuroimaging Laboratory, Santa Lucia Foundation, Rome, Italy

***Keywords: Speech emotion recognition (SER), Circumplex model of emotions, Partial least square (PLS) regression, Pearson correlation coefficient, Pitch contour characterization, Audio signal modulation.***

In the community of Human Computer Interface (HCI) researchers have been working for several years in trying to emulate a human communication system, using innovative technologies and methodologies, based on the emotion recognition in facial expressions and speech [1-3]. Speech emotion recognition (SER) [4] is a challenging framework in demanding human machine interaction systems. Standard approaches based on the categorical model of emotions reach low performance, probably due to the modelization of emotions as distinct and independent affective states. Starting from the recently investigated assumption on the dimensional circumplex model of emotions [5,6], SER systems are structured as the prediction of valence and arousal on a continuous scale in a two-dimensional domain.

In this study, we propose the use of a PLS regression model, optimized according to specific features selection procedures and trained on the Italian speech corpus EMOVO, suggesting a way to automatically label the corpus in terms of arousal and valence. In order to label the dataset according to valence and arousal, the circumplex diagram was divided into angular sectors, each having a specific angular range and a central angle ($a_k$) as shown in Fig. 1. Each emotion was also identified by the coordinates ($yv_e$, $ya_e$), e = 1, . . . ,$N_e$, with $N_e$ the number of emotions.

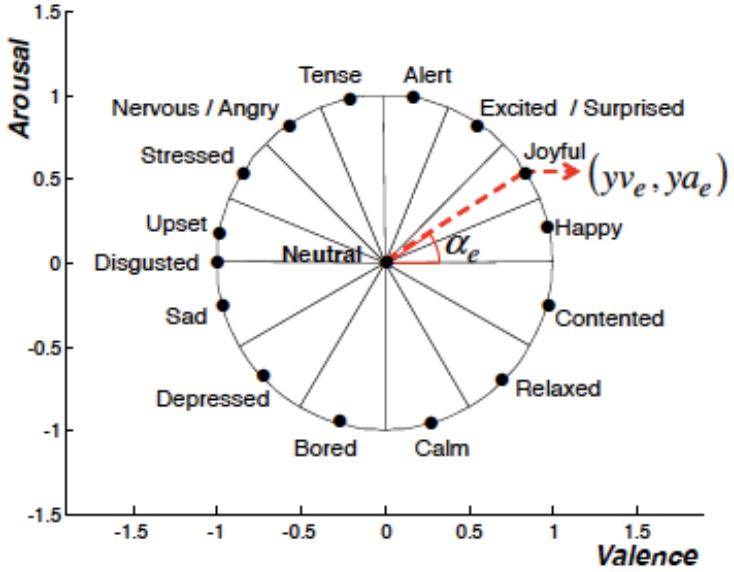


Fig. 1: The proposed circumplex diagram of emotions.

We computed 520 speech features. According to the kind of characteristics analyzed, the features can be divided into 12 different groups. New speech features related to the speech amplitude modulation, caused by the slowly-varying articulatory motion, and standard features extracted from the pitch contour, have been included in the regression model. An average value for the coefficient of determination R2 of 0:72 (maximum value of 0:95 for fear and minimum of 0:60 for sadness) is obtained for the female model and a value for R2 of 0:81 (maximum value of 0:89 for anger and minimum value of 0:71 for joy) is obtained for the male model, over the seven primary emotions (including the neutral state).

By indicating the prediction of valence and arousal on the circumplex diagram, we illustrate the two-dimensional estimation of each emotion, represented in Fig. 11(a)-11(b). Reference coordinates are indicated by the orange squared markers while the predicted output values are indicated by the small colored dots [7].

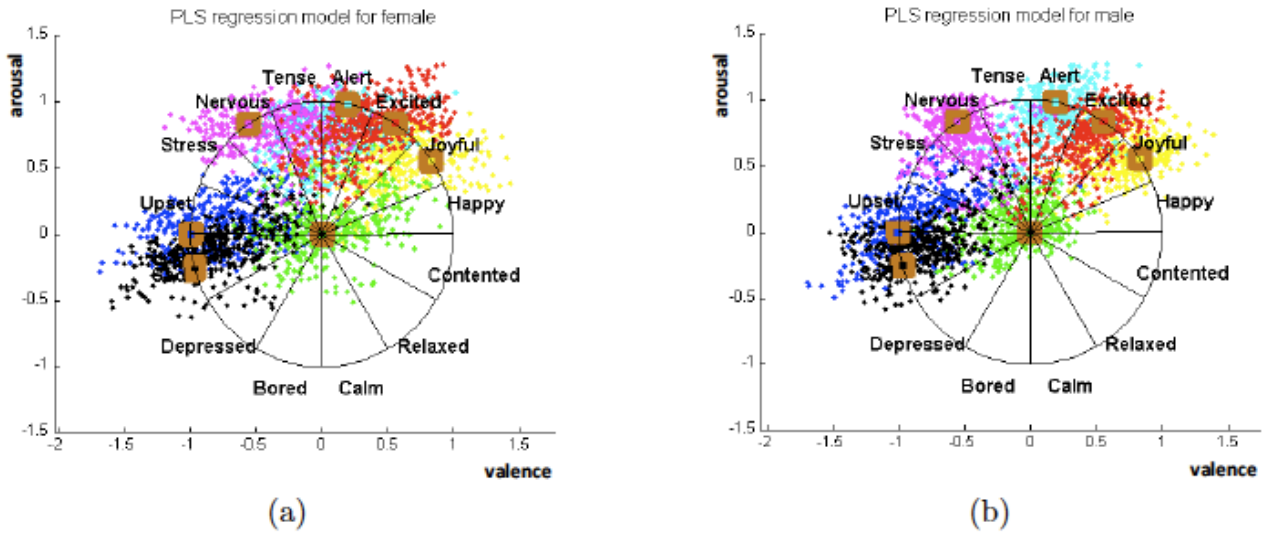


Fig. 2: Results of the PLS regression model on the circumplex diagram, using a different color for each emotion. (a) female subjects and (b) male subjects. Orange squared markers indicate the reference coordinates for each emotion.

## REFERENCES


[1] I.-O. Stathopoulou, E. Alepis, G. Tsihrintzis, M. Virvou, "On assisting a visual-facial affect recognition system with keyboard-stroke pattern information", *Knowledge-Based Systems* 23 (4) (2010) 350–356.

[2] I. Mauss, M. Robinson, "Measures of emotion: A review", *Cognition and emotion* 23 (2) (2009) 209–237.

[3] R. Nakatsu, J. Nicholson, N. Tosa, "Emotion recognition and its application to computer agents with spontaneous interactive capabilities", *Knowledge-Based Systems* 13 (7-8) (2000) 497–504.

[4] P. Ekman, W. Friesen, "Measuring facial movement", *Environmental Psychology and Nonverbal Behavior* 1 (1) (1976) 56–75.

[5] J. Russell, "A circumplex model of affect", *Journal of Personality and Social Psychology* 39 (6) (1980), pp. 1161-1178.

[6] J. Posner, J. Russell, B. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology", *Development and Psychopathology* 17 (3) (2005) 715 – 734.

[7] Mencattini A., Martinelli E., Costantini G., Todisco M., Basile B., Bozzali M., Di Natale C., (2014). "Speech emotion recognition using amplitude modulation parameters and a combined feature selection procedure", *Knowledge-Based Systems*, vol. 63, June 2014, pp. 68-81.