

Thesis Review

Affective Signal Processing (ASP): Unraveling the mystery of emotions, by Egon L. van den Broek

Anna Esposito^{a,b}

^a*Department of Psychology, Second University of Naples, Caserta, Italy*

^b*International Institute for Advanced Scientific Studies, Vietri sul Mare, Via Pellegrino 19, 84019, Salerno, Italy*
E-mail: iiass.annaesp@tin.it; anna.esposito@unina2.it

Abstract. The present work is a review of the PhD thesis defended by Egon L. van den Broek on September 16, 2011 at the department of Human Media Interaction, Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands. I was a member of the PhD dissertation Committee. I was overwhelmed by the quality and the amount of work he did. The thesis is a great contribution to our understanding and harnessing the principles underlying affective multimodal communication and to the development of future real-world technologies for affective signal processing.

Keywords: Emotions, affective interactions, emotional signal processing, emotion-aware systems, affective computing

1. Introduction

Emotion is a topic that has received much attention during the last few years, both in the context of Human Computer Interaction (HCI), which involves research themes related to speech synthesis, as well as in automatic speech recognition, interactive dialogues systems, wearable computing, embodied conversational agent systems, and intelligent avatars that are capable of performing believable actions and naturally reacting to human users [1,2]. Along these application requirements, research on emotions plays a fundamental role. In the HCI field, therefore, the research objectives are to identify methods and procedures capable of automatically identifying human emotional states exploiting the multimodal nature of emotions. This requires the consideration of several key aspects, such as the development and integration of algorithms and procedures for applications in communication, and for the recognition of emotional states, from gestures, speech, gaze, facial expressions, and bio-signals, in anticipation of the implementation of intelligent avatars and interactive dialog systems

that could be exploited to improve the learning and understanding of emotional behavior and facilitate user's access to future communication services.

The focus of this thesis was on Affective Signal Processing (ASP), and the research reported [3] provided valuable insights to our understanding and harnessing the principles underlying affective multimodal communication. In the meantime, suggesting new engineering approaches toward the improvement and development of signal processing techniques, such as combining speech and bio-signals. The underlying vision was to develop new cutting edge methodologies for recognizing emotional states. It was a huge amount of high quality research work that harnessed HCI, Artificial Intelligence (AI), and health informatics, relating the processing of bio-signals to psychological models of emotional behaviors.

The results were finely assessed at the light of the recent advances in the psychology of emotions and mathematical algorithms developed for their computational treatment. In particular, after a detailed description on the relevance of ASP for future real-world technological applications [4,5], the author

challenged two different approaches: baseline-free and bimodal ASP [4,6–8] taking into account their effectiveness in supporting the dimensional and basic psychological models of emotions [9–11]. In the light of these results, computational models of machine learning and AI were exploited and assessed to support the feasibility of emotional aware systems [12–14].

2. Baseline-free and bimodal ASP

Emotions and the related perceptual cues to infer them have always been investigated considering expressive signals such as face, voice, and physiological signals separately. Some studies sustain that facial expressions are more emotionally informative than physiological signals and vocal expressions [15], whereas others suggest that vocal expressions may signal more faithful emotions since physiological processes, such as respiration and muscle tension, are naturally influenced by emotional responses [16]. It was noted [17,18] that while emotional vocal expressions have been assessed dynamically along the time dimension because speech is intrinsically a dynamic process, for facial expressions, a long established tradition has defined emotion in terms of qualitative targets; that is, static positions capable of being displayed in a still photograph. The still image usually captures the apex of the expression; that is, the instant at which the indicators of emotion are most marked. However, in daily experience, emotional states are intrinsically dynamic processes and the associated facial expressions vary along time. These aspects were investigated in the thesis chapters dedicated to baseline free and bimodal ASP.

2.1. Baseline free analyses

The baseline free analyses showed that the skewness and kurtosis, combined with other parameters (depending on the signal), of unfiltered measurements of the Galvanic Skin Response (GSR) and of three electromyography (EMG) signals: frontalis, corrugator supercilii, and zygomaticus major discriminated among four categories of emotions: negative, positive, mixed, and neutral despite. The analyses were performed on both long (120 seconds) and short (10 seconds) time windows. The results identified generic bio-signal emotional features, prospected the development of robust signal processing algorithms for use in future emotion-aware systems, and envisioned the feasibility of emotion-aware consumer products [4] (cf. [19]).

2.2. Bi-modal analyses

The bi-modal ASP approach exploited the rare combination of speech, electrocardiogram, and a revised Self Assessment Mannequin (SAM) for assessing the influence of emotion representations, through two different emotion induction procedures. Both the dimensional valence-arousal and the six basic emotion models were utilized, taking into account the context, and personality traits such as neuroticism and extroversion. Results favored the valence-arousal model and rejected the six basic emotion model. Moreover, the results showed that the combination of Heart Rate Variability (HRV) and the fundamental frequency (F0) explained most of the participants' experienced arousal. Context and the personality trait neuroticism also played a role while extroversion seems to be of hardly any influence [6–8] (cf. [19]).

3. Toward emotion-aware computing

Nowadays, we do have emotionally colored systems, but far from human ability. The achievement of a human level machine for emotional behavior (and in general of human level automaton intelligence) raises the need for more accurate solutions to the following challenges:

- a) Identify a set of processing algorithms able to capture emotional invariant features from multimodal social signals, and in particular from speech [19];
- b) Infer simple and fast computational models trained to classify, as well as humans, emotional acoustic features for the maintenance of sentences hierarchically structured, time dependent and reciprocally connected through complex relations, such as a set of multifaceted emotional feelings [19].

The second part of the thesis surveys these topics and provides insights for a holistic approach to the analysis, recognition and synthesis of affective states.

3.1. Automatic classification of affective states

A generic framework for automated emotion classification is proposed and tested on the data collected in the previous thesis' chapters. Using these quasi-spontaneous emotional vocal expressions, it was shown that EMG and GSR parameters can adequately discriminate with up to 61.31% correct classification

between 4 emotion classes (i.e., neutral, positive, negative, and mixed), without using information on the speaker's personal profiles [14]. The results emphasize the need for a holistic approach to ASP [19,20].

3.2. Health informatics application of ASPs

This study explores the feasibility of ASP health informatics applications. Speech data from Post-Traumatic Stress Disorder (PTSD) patients were collected in a controlled storytelling (ST) and an ecologically re-living (RL) study, each consisting of a happy and an anxiety triggering session. A set of Speech features were selected (i.e., amplitude, zero crossings, power, high-frequency power, and F0) and compressed through a linear regression model and principal component analysis. The Subjective Unit of Distress (SUD) was selected as independent stress indicator and both the models were cross-validated using machine learning algorithms. It was shown that the two sessions could be discriminated in 89% (for ST) and 77% (for RL) of the cases [12,13].

4. Conclusions

This work does not provide a set of rules on how to implement intelligent emotional interfaces. Instead, it presents a personal account on how to identify a theoretical framework to extract rules from multimodal emotional data suggesting a holistic approach to the computational treatment of affective states and providing a exhaustive analysis of data and models that can help in implementing efficient Affective Signal Processing (ASP).

References

- [1] E. Aarts, Ambient intelligence: Vision of our future, *IEEE Multimedia* **11**(1) (2004), 12–19.
- [2] C. Ramos, J.C. Augusto and D. Shapiro, Ambient Intelligence – The next step for artificial intelligence, *IEEE Intelligent Systems* **23**(2) (2008), 15–18.
- [3] E.L. van den Broek, Affective Signal Processing (ASP): Unraveling the mystery of emotions. PhD thesis, Human Media Interaction (HMI), Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands, 2011.
- [4] E.L. van den Broek, Ubiquitous emotion-aware computing, *Personal and Ubiquitous Computing* **16** (2012), in press.
- [5] E.L. van den Broek, J.H. Janssen, M.D. van der Zwaag, J.H.D.M. Westerrink and J.A. Healey, Prerequisites for Affective Signal Processing (ASP) – Part V: A response to comments and suggestions, in: *Biosignals 2011: Proc. of the International Conference on Bio-inspired systems and Signal Processing*, January 26–29, Rome, Italy, F. Babiloni, A. Fred, J. Filipe and H. Gamboa, eds, 2011, pp. 301–306.
- [6] J.H.D.M. Westerink, E.L. van den Broek, M.H. Schut, J. van Herk and K. Tuinenbreijer, Computing emotion awareness through galvanic skin response and facial electromyography, in: *Probing Experience: From Assessment of User Emotions and Behaviour to Development of Products*, (Chapter 14), J.H.D.M. Westerink, M. Ouwerkerk, T. Overbeek, W.F. Pasveer and B. de Ruyter, eds, Springer Philips Research Book Series, Vol. 8, Springer Science+Business Media B.V., Dordrecht, The Netherlands, 2008, pp. 137–150.
- [7] E.L. van den Broek and J.H.D.M. Westerink, Considerations for emotion-aware consumer products, *Applied Ergonomics* **40**(6) (2009), 1055–1064.
- [8] E.L. van den Broek, M.H. Schut, J.H.D.M. Westerink and K. Tuinenbreijer, Unobtrusive Sensing of Emotions (USE), *Journal of Ambient Intelligence and Smart Environments*, **1**(3) (2009), 287–299.
- [9] J.R.J. Fontaine, K.R. Scherer, E.B. Roesch and P.C. Ellsworth, The world of emotions is not two-dimensional, *Psychological Science* **18**(12) (2007), 1050–1057.
- [10] H. Schlosberg, The description of facial expressions in terms of two dimensions, *Journal of Experimental Psychology* **44**(4) (1952), 229–237.
- [11] J.A. Russell, Culture and the categorization of emotions, *Psychological Bulletin* **110**(3) (1991), 426–450.
- [12] E.L. van den Broek, F. van der Sluis and T. Dijkstra, Telling the story and re-living the past: How speech analysis can reveal emotions in Post-Traumatic Stress Disorder (PTSD) Patients, in: *Sensing Emotions: The Impact of Context on Experience Measurements*, (Chapter 10), J.H.D.M. Westerink, M. Krans and M. Ouwerkerk, eds, Springer Philips Research Book Series, Vol. 12, Springer Science+Business Media B.V., Dordrecht, The Netherlands, 2011, pp. 153–180.
- [13] E.L. van den Broek, F. van der Sluis and T. Dijkstra, Cross-validation of bimodal health-related stress assessment, *Personal and Ubiquitous Computing* **16** (2012), in press.
- [14] E.L. van den Broek, V. Lisý, J.H. Janssen, J.D.H.M. Westerink, M.H. Schut and K. Tuinenbreijer, Affective man-machine interface: Unveiling human emotions through biosignals, in: *Biomedical Engineering Systems and Technologies: BIOSTEC2009 Selected Revised papers*, A. Fred, J. Filipe and H. Gamboa, eds, Communications in Computer and Information Science, Vol. 52, Springer-Verlag, Berlin/Heidelberg, Germany, 2010, pp. 21–47.
- [15] C.E. Izard, Innate and universal facial expressions: Evidence from developmental and cross-cultural research, *Psychological Bulletin* **115**(2) (1994), 288–299.
- [16] K. Scherer, Vocal communication of emotion: A review of research paradigms, *Speech Communication* **40**(1–2) (2003), 227–256.
- [17] A. Esposito, The perceptual and cognitive role of visual and auditory channels in conveying emotional information, *Cognitive Computation* **1**(2) (2009), 268–278.
- [18] A. Esposito, Affect in multimodal information, in: *Affective Information Processing*, (Chapter 12), J. Tao and T. Tan, eds, Springer Verlag London Limited, London, UK, 2009, 203–226.
- [19] E.L. van den Broek, Robot nannies: Future or fiction? *Interaction Studies* **11**(2) (2010), 274–282.
- [20] A. Esposito, On the recognition of emotional vocal expressions: Motivations for an holistic approach, in: *Mobile Social Signal Processing*, A. Vinciarelli, et al., eds, Springer-Verlag, Lecture Notes in Computer Science, in press.