Available online at www.sciencedirect.com
ScienceDirect
Computer Speech & Language xxx (2018) xxx-xxx
www.elsevier.com/locate/csl

# Affective and behavioural computing: Lessons learnt from the First Computational Paralinguistics Challenge

Q1 Björn Schuller[a,b,c,*], Felix Weninger[d], Yue Zhang[a,d], Fabien Ringeval[e,f], Anton Batliner[b,g], Stefan Steidl[g], Florian Eyben[e], Erik Marchi[d,e], Alessandro Vinciarelli[h], Klaus Scherer[c], Mohamed Chetouani[i], Marcello Mortillaro[c]

Q2 [a] *Imperial College London, GLAM – Group on Language, Audio, & Music, UK*
[b] *Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany*
[c] *Université de Genève, Swiss Center for Affective Sciences, Switzerland*
[d] *Machine Intelligence & Signal Processing group, Technische Universität München, Germany*
[e] *audEERING GmbH, Germany*
[f] *Université Grenoble Alpes, CNRS, Grenoble INP, LIG, Grenoble, France*
[g] *Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany*
[h] *School of Computing Science, University of Glasgow, UK*
[i] *Sorbonne Universités, Université Pierre-et-Marie-Curie, Paris, France*

Received 10 December 2016; received in revised form 1 October 2017; accepted 19 February 2018
Available online xxx

**Abstract**

In this article, we review the INTERSPEECH 2013 Computational Paralinguistics ChallengE (ComParE) – the first of its kind – in light of the recent developments in affective and behavioural computing. The impact of the first ComParE instalment is manifold: first, it featured various new recognition tasks including social signals such as laughter and fillers, conflict in dyadic group discussions, and atypical communication due to pervasive developmental disorders, as well as enacted emotion; second, it marked the onset of the ComParE, subsuming all tasks investigated hitherto within the realm of computational paralinguistics; finally, besides providing a unified test-bed under well-defined and strictly comparable conditions, we present the definite feature vector used for computation of the baselines, thus laying the foundation for a successful series of follow-up Challenges. Starting with a review of the preceding INTERSPEECH Challenges, we present the four Sub-Challenges of ComParE 2013. In particular, we provide details of the Challenge databases and a meta-analysis by conducting experiments of logistic regression on single features and evaluating the performances achieved by the participants.


*Keywords:* Computational Paralinguistics; Social Signals; Conflict; Emotion; Autism; Survey; Challenge



Q3 * Corresponding author at: Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany.
*E-mail address:* bjoern.schuller@imperial.ac.uk (B. Schuller).

http://dx.doi.org/10.1016/j.csl.2018.02.004
0885-2308/ 2018 Published by Elsevier Ltd.

JID: YCSLA [m3+;March 20, 2018;11:52]


## 1. Introduction

Affective Computing, focusing on the emotional mechanisms in natural human-machine interaction, has been an active topic for two decades now since its early emergence in the second quinquennium of the 1990s (Picard, 1997). Affective computers are aimed to recognise, express, model, communicate, and respond to emotional information, thus providing better performance in collaboration and communication with human beings (Picard, 1997). Propelled by the advances in speech processing technology, many of the suggested applications of affective computing to computer-assisted learning, perceptual information retrieval, arts and entertainment, and human health and interaction as envisioned in Picard's pioneering work have already become reality, e.g., wearable computer devices, interactive emotion games for social inclusion of people with autism spectrum condition (ASC), and big data analytic systems.

From a psychological point of view, the realm of affect extends beyond the domain of emotions and moods (Russell, 2003; Beedie et al., 2005); in current studies, the terms affect, mood, and emotion are often used interchangeably, without much effort at conceptual differentiation (Ekkekakis, 2013). In an attempt to draw some lines of demarcation, Russell (2009) advocated the concept of *core affect* as a neurophysiological state, accessible to consciousness as a simple non-reflective feeling: feeling good or bad, feeling lethargic or energised, with the two underlying dimensions of pleasure−displeasure and activation−deactivation.

Most importantly, in spite of the paramount importance of affect, it only presents one facet of human beings, thus the paradigm of affective computing has been shifting towards a more holistic understanding of human social intelligence (Albrecht, 2006). In this context, Pentland (2007) and Vinciarelli et al. (2012a) pioneered the domain of social signal processing, with the aim to endow machines with human-like emotional, social perceptual and behavioural abilities.

For speech processing, the paradigm shift has led to an increasing attention to the automatic recognition of speaker characteristics beyond affective states, which has enabled a new broad spectrum of applications such as virtual assistants with personalised aspects, safety and security monitoring services, and speaker identification systems. There is currently a wealth of loosely connected studies, mostly on affect recognition (including emotion, depression, and stress level), but also recognition of other speaker states and traits such as sleepiness, alcohol intoxication (Schiel and Heinrich, 2009), health condition (Maier et al., 2009), personality (Mohammadi et al., 2010), and biological primitives in terms of age, gender, height, weight (Krauss et al., 2002; Schuller et al., 2013). From the plethora of well studied and currently under-researched speech phenomena, a new major field of speech technology research has been emerging, termed 'computational paralinguistics' by Schuller (2012) and Schuller and Batliner (2014).

## 2. The INTERSPEECH challenges

Along with the growing maturity of this field, different research challenges have been established, allowing researchers to compare their affect recognition systems with benchmark performances, and at the same time, addressing the different channels of affect manifestations such as facial expression, body gesture, speech, and physiological signals (e.g., heart rate, skin conductivity) (Tao and Tan, 2005). For instance, the Audio/Visual Emotion Challenge and Workshop (AVEC) is aimed at bridging between different modalities by featuring audio, visual, and audiovisual analysis for spontaneous emotion recognition (Ringeval et al., 2015). Likewise, the Emotion Recognition In The Wild Challenge and Workshop (EmotiW) scopes multimodal emotion recognition, while focusing on snippets of movies (Dhall et al., 2013). The MediaEval Benchmarking Initiative for Multimedia Evaluation[1] sets a special focus on the social and human aspects of multimedia access and retrieval, while emphasising the 'multi' in multimedia involving speech recognition, content analysis, music and audio analysis, user-contributed information (tags, tweets), viewer affective response, social networks, temporal and geo-coordinates.

The INTERSPEECH Challenges 2009 to 2012 were held in conjunction with the annual INTERSPEECH conference, one of the prime venues in speech signal processing. In the following, we detail the task specifications, data, features, Challenge conditions and evaluations of this Challenge series. The first INTERSPEECH 2009 Emotion Challenge (IS09EC) (Schuller et al., 2009; 2011a) featured a binary (idle vs negative) and a five-way (anger, emphatic, neutral, positive, and rest) classification task on the FAU Aibo Emotion Corpus of naturalistic children's

[1] http://www.multimediaeval.org/.

Please cite this article as: B. Schuller et al., Affective and behavioural computing: Lessons learnt from the First Computational Paralinguistics Challenge, Computer Speech & Language (2018), http://dx.doi.org/10.1016/j.csl.2018.02.004