

## Summary for AVEC 2018: bipolar disorder and cross-cultural affect recognition

Fabien Ringeval, Björn Schuller, Michel Valstar, Roddy Cowie, Maja Pantic

### Angaben zur Veröffentlichung / Publication details:

Ringeval, Fabien, Björn Schuller, Michel Valstar, Roddy Cowie, and Maja Pantic. 2018. "Summary for AVEC 2018: bipolar disorder and cross-cultural affect recognition." In *Proceedings of the 26th ACM International Conference on Multimedia, October 22-26, 2018, Seoul, Republic of Korea*, edited by Susanne Boll, 2111–12. New York, NY: ACM.  
<https://doi.org/10.1145/3240508.3243719>.

### Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:

**Deutsches Urheberrecht**

Weitere Informationen finden Sie unter: / For more information see:

<https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/>



# Summary for AVEC 2018: Bipolar Disorder and Cross-Cultural Affect Recognition

Fabien Ringeval  
Université Grenoble Alpes, CNRS  
Grenoble, France  
fabien.ringeval@imag.fr

Björn Schuller\*  
University of Augsburg  
Augsburg, Germany  
schuller@ieee.org

Michel Valstar  
University of Nottingham  
Nottingham, UK  
michel.valstar@nottingham.ac.uk

Roddy Cowie  
Queen's University Belfast  
Belfast, UK

Maja Pantic†  
Imperial College London  
London, UK

## ABSTRACT

The eighth Audio-Visual Emotion Challenge and workshop AVEC 2018 was held in conjunction with ACM Multimedia'18. This year, the AVEC series addressed major novelties with three distinct sub-challenges: bipolar disorder classification, cross-cultural dimensional emotion recognition, and emotional label generation from individual ratings. The Bipolar Disorder Sub-challenge was based on a novel dataset of structured interviews of patients suffering from bipolar disorder (BD corpus), the Cross-cultural Emotion Sub-challenge relied on an extension of the SEWA dataset, which includes human-human interactions recorded 'in-the-wild' for the German and the Hungarian cultures, and the Gold-standard Emotion Sub-challenge was based on the RECOLA dataset, which was previously used in the AVEC series for emotion recognition. In this summary, we mainly describe participation and conditions of the AVEC Challenge.

## CCS CONCEPTS

• **General and reference** → **Performance**; • **Computing methodologies** → **Biometrics**;

## KEYWORDS

Affective Computing; Bipolar Disorder; Cross-Cultural Emotion

## ACM Reference Format:

Fabien Ringeval, Björn Schuller, Michel Valstar, Roddy Cowie, and Maja Pantic. 2018. Summary for AVEC 2018: Bipolar Disorder and Cross-Cultural Affect Recognition. In *2018 ACM Multimedia Conference (MM '18), October 22–26, 2018, Seoul, Republic of Korea*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3240508.3243719>

\*The author is further affiliated with Imperial College London, London, UK.

†The author is further affiliated with University of Twente, Twente, The Netherlands.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution.

MM '18, October 22–26, 2018, Seoul, Republic of Korea

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5665-7/18/10.

<https://doi.org/10.1145/3240508.3243719>

## 1 INTRODUCTION

This year's Audio-Visual Emotion Challenge and workshop (AVEC 2018) has been organised in conjunction with the 26th ACM International Conference on Multimedia, MM 2018, held in Seoul, Korea, 22 – 26 October 2018.

The AVEC 2018's theme is 'Bipolar Disorder and Cross-Cultural Affect' and it is the eighth competition event aimed at comparison of multimedia processing and machine learning methods for automatic audiovisual emotion and health analysis, with all of the participants competing under strictly the same conditions in this first of its kind series [4–6, 8–12]. However, further similar endeavours that have been undertaken since AVEC started in 2011 are to be noted [3].

As before, the goal of the Challenge is to compare the relative merits of the approaches for audiovisual recognition of emotion and health under well-defined and strictly comparable conditions, and establish to what extent fusion of the approaches is possible and beneficial. The main underlying motivation is the need to advance emotion recognition and health estimation, for multimedia retrieval to a level where behaviours expressed during human-human, or human-agent interactions, can be reliably sensed in real-life conditions, as this is exactly the type of data that the new generation of affect-oriented multimedia and human-machine/human-robot communication interfaces have to face in the real world.

We called for participation in three Sub-challenges: (i) Bipolar Disorder Sub-challenge (BDS): participants were required to classify audiovisual recordings of structured interviews of patients suffering from bipolar disorder – recorded periodically from the day of admittance to discharge – into *remission*, *hypo-mania*, or *mania*, as defined by the Young Mania Rating Scale [13]; (ii) Cross-cultural Emotion Sub-challenge (CES): participants were asked to predict the level of three emotional dimensions (*arousal*, *valence*, and *liking*) time-continuously from audiovisual recordings of dyadic interactions captured 'in-the-wild', i.e., recorded in various places such as home or work place, and with arbitrary personal equipment, in a cross-cultural setting with only labels of subjects with German culture available to infer the labels of subjects with Hungarian culture; (iii) Gold-standard Emotion Sub-challenge (GES): participants were required to generate a reliable gold-standard, i.e., a single time series preserving variance of emotion, from individual ratings of emotional dimensions (*arousal* and *valence*), that was then evaluated by a baseline multimodal (audio, video, physiology) emotion recognition system from recordings of dyadic interactions.

As benchmarking database, a novel dataset of bipolar disorder was used for the BDS, the BD corpus [2]. It includes audiovisual recordings of structured interviews performed by 47 Turkish speaking subjects aged 18-53. All those subjects suffered from bipolar disorder and were recruited from a mental health service hospital where they were diagnosed by clinicians following DSM-5's inclusion criteria [1]. For the CES, an extended version of the Sentiment Analysis in the Wild (SEWA) database<sup>1</sup>, with new data collected from 32 pairs of Hungarian participants aged 18-60<sup>+</sup> in the same conditions as for the 32 others pairs of German participants with same age range, was used as a blind test set for the first ever cross-cultural (German → Hungarian) emotion recognition competition task. For the GES, we used the RECOLA dataset [7], which contains audiovisual and physiological recordings of dyadic interactions from 27 French speaking subjects aged 18-25.

Besides participation in the Challenge we called for papers addressing the overall topics of this workshop. In the following sections, we describe the participation in this year and outline the conditions for participation in particular in the competitive challenge event. We further acknowledge those that helped realise AVEC 2018.

## 2 CHALLENGE CONDITIONS

As in previous years, we required to sign an end user license agreement to access the data. After downloading the data, participants could directly start their own experiments with the train and development sets. In addition, standard feature sets were provided for audiovisual data, along with scripts available in a public repository<sup>2</sup>, which participants were free to use for reproducing both the baseline features and recognition systems. Once they found their best method they had to write a paper for the workshop. At the same time, they could compute their results per instance of the test set. Participants' results needed to be sent as a single packed file per Sub-challenge to the organisers by email and scores were returned within 24 hours during typical working days. Each participant had up to five submission attempts per Sub-challenge.

## 3 PARTICIPATION

The call for participation and papers attracted registrations of 57 teams from all over the world, with 41, 44, and 45 teams participating in the BDS, CES, and GES, respectively. 11 teams submitted results for the BDS, 8 teams for the CES, and 4 teams for the GES. Finally, 23 paper submissions were received, which were assigned three reviewers, each, and reviewed independently. AVEC 2018 reviewing was double blind, and acceptance was based on relevance to the workshop, novelty, technical quality, and performance on the test partition. The program committee accepted 11 papers in addition to the independently reviewed baseline paper as oral presentation. Again, we hope that these proceedings will serve as a valuable reference for researchers and developers in the area of audiovisual emotion recognition and health analysis in real-life settings.

<sup>1</sup><http://sewaproject.eu>

<sup>2</sup><https://github.com/AudioVisualEmotionChallenge/AVEC2018>

## ACKNOWLEDGMENTS

The research leading to these results has received funding from the Horizon 2020 Programme through the Innovative Action No. 645094 (SEWA), and the Research Innovative Action No. 645378 (ARIA-VALUSPA), and No. 688835 (DE-ENIGMA), and from the University of Fribourg, Switzerland, through the commercial licenses of the RECOLA project. The authors further thank the sponsors of the challenge – audeERING GmbH and the Association for the Advancement of Affective Computing (AAAC) – and the technical program committee members for the quality of their review.

## REFERENCES

- [1] 2013. Diagnostic and Statistical Manual of mental disorders (5th Ed.). American Psychiatric Association, Washington, DC.
- [2] Elvan Çiftçi, Heysem Kaya, Hüseyin Güleç, and Albert Ali Salah. 2018. The Turkish audio-visual Bipolar Disorder corpus. In *Proceedings of the 1st Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*. AAAC, Beijing, China.
- [3] Abhinav Dhall, Roland Göcke, Jyoti Joshi, Michael Wagner, and Tom Gedeon. 2013. Emotion recognition in the wild challenge (EmotiW) challenge and workshop summary. In *Proceedings of the 15th ACM International Conference on Multimodal Interaction, ICMi 2013*. ACM, 371–372.
- [4] Fabien Ringeval, Björn Schuller, Michel Valstar, Roddy Cowie, Heysem Kaya, Maximilian Schmitt, Shahin Amiriparian, Nicholas Cummins, Dennis Lalanne, Adrien Michaud, Elvan Ciftci, Hüseyin Güleç, Albert Ali Salah, and Maja Pantic. 2018. AVEC 2018 Workshop and Challenge: Bipolar Disorder and Cross-Cultural Affect Recognition. In *Proceedings of the 8th International Workshop on Audio/Visual Emotion Challenge (AVEC)*, co-located with the 26th ACM International Conference on Multimedia (ACM MM). ACM, Seoul, South Korea. 11 pages, to appear.
- [5] Fabien Ringeval, Björn Schuller, Michel Valstar, Jonathan Gratch, Roddy Cowie, Stefan Scherer, Sharon Mozgai, Nicholas Cummins, and Maja Pantic. 2017. AVEC 2017 – Real-life depression, and affect recognition workshop and challenge. In *Proceedings of the 7th International Workshop on Audio/Visual Emotion Challenge (AVEC)*, co-located with the 25th ACM International Conference on Multimedia (ACM MM). ACM, Mountain View, CA, USA, 3–9.
- [6] Fabien Ringeval, Björn Schuller, Michel Valstar, Shashank Jaiswal, Erik Marchi, Denis Lalanne, Roddy Cowie, and Maja Pantic. 2015. AV+EC 2015 – The first affect recognition challenge bridging across audio, video, and physiological data. In *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge (AVEC)*, co-located with the ACM International Conference on Multimedia (ACM MM). ACM, Brisbane, Australia, 3–8.
- [7] Fabien Ringeval, Andreas Sonderegger, Jürgen Sauer, and Denis Lalanne. 2013. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In *Proceedings of the 2nd International Workshop on Emotion Representation, Analysis and Synthesis in Continuous Time and Space (EmoSPACE)*, held in conjunction with the 10th International IEEE Conference on Automatic Face and Gesture Recognition (FG). IEEE, Shanghai, China. 8 pages.
- [8] Björn Schuller, Michel Valstar, Florian Eyben, Roddy Cowie, and Maja Pantic. 2012. AVEC 2012 – The continuous Audio/Visual Emotion Challenge. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI)*. ACM, Santa Monica, CA, USA, 449–456.
- [9] Björn Schuller, Michel Valstar, Florian Eyben, Gary McKeown, Roddy Cowie, and Maja Pantic. 2011. AVEC 2011 – The First International Audio/Visual Emotion Challenge. In *Proceedings of the 4th biannual International Conference on Affective Computing and Intelligent Interaction (ACII)*, Vol. II. Springer, Memphis, TN, USA, 415–424.
- [10] Michel Valstar, Jonathan Gratch, Björn Schuller, Fabien Ringeval, Roddy Cowie, and Maja Pantic. 2016. Summary for AVEC 2016: Depression, mood, and emotion recognition workshop and challenge. In *Proceedings of the 24th ACM International Conference on Multimedia (ACM MM)*. ACM, Amsterdam, The Netherlands, 1483–1484.
- [11] Michel Valstar, Björn Schuller, Jarek Krajewski, Roddy Cowie, and Maja Pantic. 2013. Workshop summary for the 3rd international Audio/Visual Emotion Challenge and workshop. In *Proceedings of the 21st ACM International Conference on Multimedia (ACM MM)*. ACM, Barcelona, Spain, 1085–1086.
- [12] Michel Valstar, Björn Schuller, Jarek Krajewski, Roddy Cowie, and Maja Pantic. 2014. AVEC 2014: The 4th international Audio/Visual Emotion Challenge and workshop. In *Proceedings of the 22nd ACM International Conference on Multimedia (ACM MM)*. ACM, Orlando, FL, USA, 1243–1244.
- [13] Robert C. Young, Jeffery T. Biggs, Veronika E. Ziegler, and Dolores A. Meyer. 1978. A rating scale for mania: Reliability, validity and sensitivity. *The British Journal of Psychiatry* 133, 5 (November 1978), 429–435.