

AVEC'19: audio/visual emotion challenge and workshop

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

Angaben zur Veröffentlichung / Publication details:

Ringeval, Fabien, Björn Schuller, Michel Valstar, Nicholas Cummins, Roddy Cowie, and Maja Pantic. 2019. "AVEC'19: audio/visual emotion challenge and workshop." In *Proceedings of the 27th ACM International Conference on Multimedia - MM '19, October 21 - 25, 2019, Nice, France*, 2718–19. New York, NY: ACM Press.
<https://doi.org/10.1145/3343031.3350550>.

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/>



AVEC'19: Audio/Visual Emotion Challenge and Workshop

Fabien Ringeval
Université Grenoble Alpes, CNRS
Grenoble, France

Björn Schuller*
University of Augsburg
Augsburg, Germany

Michel Valstar
University of Nottingham
Nottingham, UK

Nicholas Cummins
University of Augsburg
Augsburg, Germany

Roddy Cowie
Queen's University Belfast
Belfast, UK

Maja Pantic†
Imperial College London
London, UK

ABSTRACT

The ninth Audio-Visual Emotion Challenge and workshop AVEC 2019 was held in conjunction with ACM Multimedia'19. This year, the AVEC series addressed major novelties with three distinct tasks: State-of-Mind Sub-challenge (SoMS), Detecting Depression with Artificial Intelligence Sub-challenge (DDS), and Cross-cultural Emotion Sub-challenge (CES). The SoMS was based on a novel dataset (USoM corpus) that includes self-reported mood (10-point Likert scale) after the narrative of personal stories (two positive and two negative). The DDS was based on a large extension of the DAIC-WOZ corpus (c.f. AVEC 2016) that includes new recordings of patients suffering from depression with the virtual agent conducting the interview being, this time, wholly driven by AI, i. e., without any human intervention. The CES was based on the SEWA dataset (c. f. AVEC 2018) that has been extended with the inclusion of new participants in order to investigate how emotion knowledge of Western European cultures (German, Hungarian) can be transferred to the Chinese culture. In this summary, we mainly describe participation and conditions of the AVEC Challenge.

CCS CONCEPTS

• General and reference → Performance;

KEYWORDS

Affective Computing; State-of-Mind; Cross-Cultural Emotion

ACM Reference Format:

Fabien Ringeval, Björn Schuller, Michel Valstar, Nicholas Cummins, Roddy Cowie, and Maja Pantic. 2019. Summary for AVEC 2019: State-of-Mind, Detecting Depression with AI, and Cross-Cultural Affect. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19)*, October 21–25, 2019, Nice, France. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3343031.3350550>

*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 '19, October 21–25, 2019, Nice, France

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6889-6/19/10.

<https://doi.org/10.1145/3343031.3350550>

1 INTRODUCTION

This year's Audio-Visual Emotion Challenge and workshop (AVEC 2019) has been organised in conjunction with the 27th ACM International Conference on Multimedia, MM 2019, held in Nice, France, 21 – 25 October 2019.

The AVEC 2019's theme is 'State-of-Mind, Detecting Depression with AI, and Cross-Cultural Affect Recognition' and it is the ninth 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 [2, 5–12].

As before, the goal of the Challenge is to advance health and emotion recognition systems by providing a common benchmark test set for multimodal information processing, in order to compare the relative merits of the approaches to automatic health and emotion analysis under well-defined conditions, with large volumes of un-segmented, non-prototypical and non-preselected data of wholly naturalistic behaviour. These conditions are precisely those 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) State-of-Mind Sub-challenge (SoMS): participants were required to predict the level of self-reported mood (10-point Likert scale) after the narrative of personal stories (two positive and two negative); (ii) Detecting Depression with AI Sub-challenge (DDS): participants needed to infer the level of depression severity from recordings of patients that were interacting with a virtual agent conducting automatically a clinical interview; (iii) 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 and Hungarian culture available to infer the labels of subjects with Chinese culture.

As benchmarking database, the Ulm state of mind (USoM) corpus [4] was used to assess automatically the current state of mind (SOM) from personal storytelling; such data-driven recognition system has the potential to assist the society in various ways [2]. The USoM dataset includes two positive and negative stories each told by 111 subjects for a total duration of more than 34 hours.

The Extended Distress Analysis Interview Corpus (E-DAIC) [1] was used to predict the level of depression from semi-clinical interviews conducted by an animated virtual interviewer, which was either controlled by a human (wizard) in another room for the WoZ interviews, or acting in a fully autonomous way, using different automated perception and behaviour generation modules, for the AI interviews. Overall, the E-DAIC includes interviews of 275 subjects for a total duration of more than 73 hours.

For the CES, an extended version of the Sentiment Analysis in the Wild (SEWA) database [3], with new data collected from Chinese participants aged 18–60⁺ in the same conditions as for the others pairs of German and Hungarian participants with same age range, was used as a blind test set for the cross-cultural time-continuous emotion recognition task, with the aim to investigate how emotion knowledge of Western European cultures (German, Hungarian) can be transferred to the Chinese culture. Overall, the SEWA dataset includes chats from 200 subjects for a total duration of more than 8 hours.

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 2019.

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 experiments with the train and development sets. In addition, standard feature sets were provided for audiovisual data, along with scripts available in a public repository¹, 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 61 teams from all over the world, with 36, 38, and 48 teams participating in the SoMS, CES, and DDS, respectively. 10 teams submitted results for the DDS, 7 teams for the CES, and 3 teams for the SoMS. Finally, 19 paper submissions were received, which were assigned three reviewers, each, and reviewed independently following the ACM MM reviewing guidelines. AVEC 2019 acceptance was based on relevance to the workshop, novelty, technical quality, and performance on the test partition. The program committee accepted 9 papers in addition to the independently reviewed baseline paper as an 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.

¹<https://github.com/AudioVisualEmotionChallenge/AVEC2019>

ACKNOWLEDGMENTS

The research leading to these results has received funding from the Horizon 2020 Programme through the Research Innovation Action No. 826506 (sustAGE), and No. 688835 (DE-ENIGMA). Further funding has also been received from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No. 115902, which receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. The authors further thank the sponsor of the challenge – audeERING GmbH, and the technical program committee members² for the quality of their review.

REFERENCES

- [1] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jonathan Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Alberto Rizzo, and Louis-Philippe Morency. 2014. SimSensei kiosk: A virtual human interviewer for healthcare decision support. In *Proc. of AAMAS'14*. ACM, Paris, France, 1061–1068.
- [2] Fabien Ringeval and Björn Schuller and Michel Valstar and Nicholas Cummins and Roddy Cowie and Leili Tavabi and Maximilian Schmitt and Sina Alisamir and Shahin Amiriparian and Eva-Maria Messner and Siyang Song and Shuo Lui and Ziping Zhao and Adria Mallol-Ragolta and Zhao Ren, and Mohammad Soleymani, and Maja Pantic. 2019. AVEC 2019 Workshop and Challenge: State-of-Mind, Depression with AI, and Cross-Cultural Affect Recognition. In *Proc. of AVEC'19, co-located with ACM MM 2019*. ACM, Nice, France.
- [3] Jean Kossaifi, Robert Walecki, Yannis Panagakis, Jie Shen, Maximilian Schmitt, Fabien Ringeval, Jing Han, Vedhas Pandit, Björn Schuller, Kam Star, Elnar Hajiyeve, and Maja Pantic. 2019. SEWA DB: A Rich Database for Audio-Visual Emotion and Sentiment Research in the Wild. <https://arxiv.org/abs/1901.02839>. 17 pages.
- [4] Eva-Maria Rathner, Yannik Terhorst, Nicholas Cummins, Björn Schuller, and Harald Baumeister. 2018. State of Mind: Classification through Self-reported Affect and Word Use in Speech. In *Proc. of INTERSPEECH 2018*. ISCA, Hyderabad, India, 267–271.
- [5] Fabien Ringeval, Björn Schuller, Michel Valstar, Roddy Cowie, and Maja Pantic. 2015. AVEC 2015 – The 5th International Audio/Visual Emotion Challenge and Workshop. In *Proc. of ACM MM 2015*. ACM, Brisbane, Australia, 1335–1336.
- [6] Fabien Ringeval, Björn Schuller, Michel Valstar, Roddy Cowie, and Maja Pantic. 2017. Summary for AVEC 2017 – Real-life depression and affect challenge and workshop. In *Proc. of ACM MM 2017*. ACM, Mountain View, CA, USA, 1963–1964.
- [7] Fabien Ringeval, Björn Schuller, Michel Valstar, Roddy Cowie, and Maja Pantic. 2018. Summary for AVEC 2018: Bipolar disorder and cross-cultural affect recognition. In *Proc. of ACM MM 2018*. ACM, Seoul, Republic of Korea, 2111–2112.
- [8] Björn Schuller, Michel Valstar, Florian Eyben, Roddy Cowie, and Maja Pantic. 2012. AVEC 2012 – The continuous Audio/Visual Emotion Challenge. In *Proc. of ICMI 2012*. 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 *Proc. of ACII 2011*, 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 *Proc. of ACM MM 2016*. 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 *Proc. of ACM MM 2013*. 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 *Proc. of ACM MM 2014*. ACM, Orlando, FL, USA, 1243–1244.

²The AVEC'19 Technical Program Committee was composed of Abhinav Dhall, Adria Mallol-Ragolta, Anna Esposito, Arianna Mencattini, Chi-Chun (Jeremy) Lee, Chloé Clavel, Dominique Vaufray, Dongmei Jiang, Elisabeth André, Emily Mower Provost, Felix Burkhardt, Florian Eyben, Hayley Hung, Heidi Christensen, Heysem Kaya, Jarek Krajewski, Jing Han, Julien Epps, Jun Deng, Kevin Bailly, Lang He, Lionel Prevost, Mathew Magimai Doss, Mohamed Chetouani, Rama Chellappa, Roland Göcke, Shri Narayanan, Shuo Liu, Sidney D'Mello, Simone Hantke, Theodora Chaspari, Ting (Sally) Dang, Vidhyasaharan Sethu, Zhao Ren, Zhaocheng (David) Huang, and Zixing Zhang.