The AffectMove 2021 Challenge - Affect Recognition from Naturalistic Movement Data

Temitayo Olugbade
University College London
London, United Kingdom
temitayo.olugbade.13@ucl.ac.uk

Roberto Sagoleo
Università di Genova
Genova, Italy

Simone Ghisio
Università di Genova
Genova, Italy

Nicolas Gold
University College London
London, United Kingdom

Amanda C de C Williams
University College London
London, United Kingdom

Beatrice de Gelder
Maastricht University
Maastricht, The Netherlands

Antonio Camurri
Università di Genova
Genova, Italy

Gualtiero Volpe
Università di Genova
Genova, Italy

Nadia Bianchi-Berthouze
University College London
London, United Kingdom

***Abstract*—We ran the first Affective Movement Recognition (AffectMove) challenge that brings together datasets of affective bodily behaviour across different real-life applications to foster work in this area. Research on automatic detection of naturalistic affective body expressions is still lagging behind detection based on other modalities whereas movement behaviour modelling is a very interesting and very relevant research problem for the affective computing community. The AffectMove challenge aimed to take advantage of existing body movement datasets to address key research problems of automatic recognition of naturalistic and complex affective behaviour from this type of data. Participating teams competed to solve at least one of three tasks based on datasets of different sensors types and real-life problems: multimodal EmoPain dataset for chronic pain physical rehabilitation context, weDraw-1 Movement dataset for maths problem solving settings, and multimodal Unige-Maastricht Dance dataset. To foster work across datasets, we also challenged participants to take advantage of the data across datasets to improve performances and also test the generalization of their approach across different applications.**

***Index Terms*—affective computing, bodily, challenge, datasets, movement**

I. Introduction

The AffectMove 2021 challenge brought together affective body movement datasets from very different contexts with the aim of promoting research in affect recognition from movement data and pushing the bounds on the value of such technology in real applications. This is a valuable endeavour given that body movement is a fundamental component of everyday living both in the execution of the actions that make up physical functioning as well as in rich expression of affect, cognition, and intent [1]–[4]. Yet, despite the plethora of evidence that highlight the primacy of understanding body movement for rewarding interactions with humans [5], [6] and findings of the possibility of automatic detection of affect/cognition/intent from body movement data [7], there are few technologies that currently do these in the real world.

This work was supported by the EU Future and Emerging Technologies Proactive Programme H2020 (Grant No. 824160: EnTimeMent - https://entimement.dibris.unige.it/).

The challenge addresses this state of affairs first with datasets representative of unconstrained everyday settings. In such settings, affective/cognitive experiences are spontaneous and bodily expressions of these experiences manifest as a modulation of movement execution during physical activity rather than simply being a distinct or isolated gesture [8]. Further, in real-life settings, where there are no clear breaks between activities, data is captured continuously over multiple activity types without manual activity segmentation. Additionally, obtaining training data is not trivial in real world settings resulting in the need to address such problems using relatively limited training data. Secondly, the challenge is based on data from multiple contexts and population groups captured using different sets of movement sensors. The significance of having different contexts is the divergent movement repertoires and affective/cognitive experiences across these contexts. For example, the types of movements and experiences that are important to address in everyday movements at home for people with chronic pain are not the same as the moves and expertise expressed in professional dance. This reflects one of the barriers to making affect recognition technology widely available in the real world, and so, one of the aims of the challenge was to drive the exploration of systems (architectures and/or models) that could work across contexts.

The AffectMove 2021 challenge was made up of three tasks based on three body movement datasets respectively: EmoPain [9], weDraw-1 Movement [10], and Unige-Maastricht Dance [11]–[13] datasets. All three datasets were built on deep understanding of the requirements of automatic detection technology for the given context and so the affective/cognitive experiences that they capture are application-specific states rather than the so-called basic emotions that are traditionally explored. Challenge participants could choose to work on one or more tasks of the challenge. Data for each task was split into training, validation, and test partitions. All three partitions were available to participants but they did not have access to the ground truth labels for the test partition. The predictions for each task were evaluated against the unseen ground truth labels for the test set using F1 score per label class, Matthews Correlation Coefficient [14], and accuracy. The participating team with a set of complete predictions for a given task and with the highest performance based on these metrics was selected as the winner for the task.

In the rest of this paper, we review the state of the art in the areas of bodily-expressed affect recognition in the context of pain, learning, and dance, which are relevant to the challenge. We then provide a description of the challenge tasks and data.

II. Related Work: Automatic Detection Based on Bodily Behaviour

In this section, we review studies in the area of automatic detection of affective bodily expression within the three contexts covered in the AffectMove 2021 challenge: pain, learning, and dancing.

A. Pain

Most studies on automatic detection of behaviour associated with spontaneous pain (as opposed to experimentally-induced pain, which is momentary and less threatening) have focused on facial expressions. Nevertheless, there are a number of studies that have been based on bodily expressions and/or movement features. One of the more common applications in this area is the detection of acute pain in infants in hospital settings. For example, [15] used video data in discriminating between two pain levels a couple of hours after surgery. They obtained accuracy of 0.67 based on bodily features alone and accuracy of 0.79 including facial and audio features, using hold-out validation.

Another typical use case is the discrimination between people with and without a specific chronic pain condition (e.g. Complex Regional Pain Syndrome [16], knee osteoarthritis [17]). [17] went on further to predict both current pain severity (less severe versus severe) and long-term clinical outcome (positive or not) based on gait features during stairs ascent. They achieved 0.83 and 0.70 accuracies respectively with leave-one-out cross-validation. [18] similarly used lower back muscle activity data during trunk movements such as flexion-extension to predict the outcome (positive or not) of a physical rehabilitation program for people having low back pain with 0.97 accuracy.

A related use scenario is the automatic detection of protective behaviour, i.e. behaviour intended to protect from harm or pain exacerbation [19], in people with chronic musculoskeletal pain. Studies on this topic have been based on the EmoPain dataset [9] which consists of motion capture and muscle activity data of people with and without chronic musculoskeletal pain captured during movements typical of everyday functioning, e.g. sit-to-stand. The dataset was previously released within the ambit of the EmoPain 2020 challenge [20], as part of the bid to advance the state of the art for this use case. [21] obtained the highest performance of 0.94 accuracy based on hold-out validation with unseen subjects in the test set in that competition. Beyond this, [22] even more recently achieved accuracy of 0.88 accuracy with leave-one-subject-out cross-validation in further investigation based on the motion capture data alone.

The AffectMove 2021 challenge aims to further advance the progress made in the area by employing data from the people with chronic musculoskeletal pain only, i.e. not including data from the healthy control participants, for automatic protective behaviour detection. This subset of the EmoPain dataset is more pertinent to building technology that detects protective behaviours, while a person with chronic musculoskeletal pain performs harmless physical activities, so as to provide timely, personalised feedback and/or support to help the person execute movements more fluidly.

B. Learning

In learning applications where bodily expression are considered, the focus is usually on upper body behaviour. This is perhaps due to the typical use of camera-based sensors as well as the seated settings of conventional learning environments, e.g. serious games [23] or online learning both based on a desktop computer [24], [25], traditional style classroom lectures [26], [27]. Although a wide variety of experience (such as boredom, confusion, delight, and frustration [26], [27]) has been investigated in this area, engagement may be one of the more studied states. In [25], 0.58 accuracy with hold-out validation was achieved for discrimination between four levels of engagement in young adults during online learning sessions. When features of facial expression were used in addition to the upper body features, the authors obtained accuracy of 0.61.

In [28] where the body itself was leveraged as a learning tool, the authors found performance of 0.63 F1 score (with 10-fold cross-validation) in discriminating between three levels of engagement in children using full-body features. The setting used in their study is similar to that investigated in [10] based on the weDraw-1 Movement dataset which consists of motion capture data captured while children explored mathematical problems using their bodies, e.g. in full-body rotation as dynamic representation of an angle. The weDraw-1 Movement dataset [10] further includes annotations of observed reflective thinking. The authors obtained 0.79 accuracy (average F1 score = 0.79) using hold-out validation in preliminary experiments on automatic recognition of time periods labelled as moments when reflective thinking was observed. In the AffectMove 2021 challenge, reflective thinking detection based on continuous window segmentation is pursued to extend the state of the art beyond the manual segmentation, based on the known onset and offset of observed reflective thinking periods, used in [10].

C. Dancing

Body movement is intuitively central to dance and has been the primary modality used in computational analysis of dance movements. For example, [29] used data captured with marker-based optical motion capture sensors for automatic classification of skill levels of salsa dancers into 3: beginner, intermediate, advanced. The authors obtained 0.81 and 0.64 accuracies for basic and improvised salsa steps respectively. Personal dance style has also been automatically detected from optical motion capture data captured during hip hop dances with 0.99 accuracy for beginner dancers and 0.92 accuracy for more experienced dancers [30]. The results in [30] were based on 12 dancers in either group of dancers (beginner and experienced) and hold-out validation with unseen dance instances in the test set.

More pertinent to automatic recognition of affective expressions are studies such as [31] which explored differentiation between fear, anger, grief, and joy dance expressions using body movement features extracted from video data. More recent work in [12] similarly investigated automatic detection of levels of lightness and fragility as dance movement expressions. This work was based on the Unige-Maastricht Dance dataset [11]–[13] made up of inertia sensor, muscle activity, video, and audio data captured from dancers with multiple dance backgrounds and years of experience while they performed improvised dance choreography. 0.86 accuracy was obtained for two-level classification of lightness while 0.77 accuracy was obtained for fragility, both based on leave-one-out cross-validation.

III. Challenge Description

This section describes each of the three tasks of the AffectMove 2021 challenge.

A. *Protective Behaviour Detection Based on Multimodal Body Movement Data*

The aim of this task was to advance the state of the art in automatic detection of protective behaviours in continuous streams of body movement data of people with chronic musculoskeletal pain. This contributes to the long-term goal of addressing chronic pain, which is a major healthcare challenge [32]. We envision that technology that is able to assess protective behaviour could support the delivery of personalised therapies [33], [34] in the management of the condition for the purpose of improving engagement in valued everyday activities.

For this task, we provided anonymised three-dimensional full-body joint positions and concomitant back muscle activity data for 19 people with chronic low back pain, from the EmoPain dataset [9]. The data was given in prepared three-second windows (the sampling rate of the data was 60 Hertz) and with the corresponding protective behaviour label (*present* or *absent*) for each window. A window was labelled with protective behaviour *present* if 50% of the window was rated as showing at least 1 one protective behaviour by 2 or more of four clinician raters and as protective behaviour *absent* otherwise. The windows are based on continuous segmentation of the data in which the subject went from one activity type (an exercise movement such as sit-to-stand) to the other, and so a window could include multiple activity types. For example, the conclusion of a sit-to-stand and the beginning of a return to seated position could occur within the same window.

We included the frame-level activity labels for each window to allow the use of the activity labels in predicting protective behaviour although we specified that it must not be used as an input feature. The primary reason for this requirement is that in real-life application of protective behaviour detection technology, the type of activity being performed is likely to be unknown and so unavailable as an input feature. The activity labels could instead themselves be taken as labels to be predicted, e.g., in an hierarchical architecture where activity class is predicted and then the prediction used as input for protective behaviour prediction or in a multitask framework where activity and protective behaviour are separately but simultaneously learnt.

We provided training, validation, and test sets which comprised 5,827, 1,844, and 2,744 windows from 10, 4, and 5 people with chronic musculoskeletal pain respectively.

B. *Detection of Reflective Thinking Based on Body Movement Data*

The aim of this task was to pioneer continuous detection of reflective thinking in children during mathematical problem-solving activities. Understanding mathematical ideas such as angles and shapes is a key part of basic education. Thus, digital learning technology that promotes the use of body movement as well as further recognizes critical learning moments (e.g. reflective thinking) could support learning of abstract mathematical ideas which may otherwise be challenging to relate to [35].

For this task, we provided anonymised three-dimensional full-body joint positions for 24 children from the weDraw-1 Movement dataset [10]. The data was provided in five-second windows (the sampling rate of the data was 30 Hertz) with the corresponding reflective thinking label (*observed* or *not observed*). A window was labelled with reflective thinking *observed* if 50% of the window was rated as showing reflective thinking behaviour and as reflective thinking *not observed* otherwise. We additionally included the label of the maths problem being explored in each window. As with the protective behaviour detection task, we specified that this label must not be used as an input feature.

The training, validation, and test sets for this task consisted of 2,090, 792, and 672 windows from 13, 5, and 6 children respectively.

C. *Detection of Lightness and Fragility in Dance Movement Based on Multimodal Data*

The aim of this task was to further develop the state of the art in the detection of lightness and fragility [11], [12] in dance movement. Automatic detection of such qualities is valuable for informing interactive sonification for the purpose of enriching audience experience of a dance performance as well as for training purposes to help dancers improve their skills [12].

For this task, we provided wrists, ankles, and waist accelerometer, video (with faces blurred), and audio respiration data for 13 dancers from the Unige-Maastricht Dance dataset [11]–[13]. The data included corresponding labels for levels of lightness and fragility. These labels were based on observer annotations provided by experts [12] and non-experts [13].

IV. Conclusion

We recruited participants using mailing lists of relevant conferences and related communities as well as via social media. The challenge ran between March and June 2021 and concluded with five participating teams across three continents.

The performance of the predictions submitted by these teams will be revealed at the Affective Movement Recognition Workshop of the Affective Computing and Intelligent Interaction conference. Full details of their methods and results are published in the same proceedings as the current paper.

References

[1] Sarah-Jayne, Blakemore and Jean Decety. 2001. “From the Perception of Action to the Understanding of Intention.” Nature Reviews Neuroscience 2 (August): 561–67. https://doi.org/10.1038/35086023

[2] Beatrice de Gelder. 2009. “Why Bodies? Twelve Reasons for Including Bodily Expressions in Affective Neuroscience.” Philosophical Transactions of the Royal Society B: Biological Sciences 364 (1535): 3475–84.

[3] Mijung Kim, Wolff-michael Roth, and Jennifer Thom. 2011. “Children’s Gestures and the Embodied Knowledge of Geometry.” International Journal of Science and Mathematics Education 9 (1): 207–38.

[4] Giovanni Pezzulo, Francesco Donnarumma, Haris Dindo, Alessandro D’Ausilio, Ivana Konvalinka, and Cristiano Castelfranchi. ”The body talks: Sensorimotor communication and its brain and kinematic signatures.” Physics of life reviews 28 (2019): 1-21.

[5] Gergely Csibra and György Gergely. 2007. “’Obsessed with Goals ’: Functions and Mechanisms of Teleological Interpretation of Actions in Humans .” Acta Psychologica 124: 60–78.

[6] Manuel Varlet, Ludovic Marin, Julien Lagarde, and Benoît G. Bardy. ”Social postural coordination.” Journal of Experimental Psychology: Human Perception and Performance 37, no. 2 (2011): 473.

[7] Fatemeh Noroozi, Dorota Kaminska, Ciprian Corneanu, Tomasz Sapinski, Sergio Escalera, and Gholamreza Anbarjafari. 2018. “Survey on Emotional Body Gesture Recognition.” IEEE Transactions on Affective Computing. https://doi.org/10.1109/TAFFC.2018.2874986.

[8] Michelle Karg, Ali-Akbar Samadani, Rob Gorbet, Kolja Kühnlenz, Jesse Hoey, and Dana Kulić. ”Body movements for affective expression: A survey of automatic recognition and generation.” IEEE Transactions on Affective Computing 4, no. 4 (2013): 341-359.

[9] Aung, Min S H, Sebastian Kaltwang, Bernardino Romera-paredes, Brais Martinez, Matteo Cella, Michel Valstar, Hongying Meng, et al. 2016. “The Automatic Detection of Chronic Pain- Related Expression : Requirements , Challenges and a Multimodal Dataset.” IEEE Transactions on Affective Computing 7 (4): 1–18.

[10] Olugbade, Temitayo A, Joseph Newbold, Rose Johnson, Erica Volta, Paolo Alborno, Max Dillon, Gualtiero Volpe, and Nadia Bianchi-berthouze. 2020. “Automatic Detection of Reflective Thinking in Mathematical Problem Solving Based on Unconstrained Bodily Exploration.” IEEE Transactions on Affective Computing.

[11] A. Camurri, G. Volpe, S. Piana, M. Mancini, R. Niewiadomski, N. Ferrari, C. Canepa. “The Dancer in the Eye: Towards a Multi-Layered Computational Framework of Qualities in Movement”, Proceedings of the 3rd International Symposium on Movement and Computing; 2016: ACM. p. 6.

[12] R.Niewiadomski, M. Mancini, S. Piana, P. Alborno, G. Volpe, A. Camurri. “Low-intrusive recognition of expressive movement qualities”, Proceedings of the 19th ACM International Conference on Multimodal Interaction; 2017: ACM. p. 230–237.

[13] M.J. Vaessen, E. Abassi, M. Mancini, A. Camurri, B. de Gelder. 2018. ”Computational feature analysis of body movements reveals hierarchical brain organization.” Cerebral Cortex 29 (8): 3551-3560.

[14] Matthews, B. W. 1975. “Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme.” Biochimica et Biophysica Acta 405 (2): 442–51. https://doi.org/10.1016/0005-2795(75)90109-9.

[15] Md Sirajus Salekin, Ghada Zamzmi, Dmitry Goldgof, Rangachar Kasturi, Thao Ho, and Yu Sun. 2021. “Multimodal Spatio-Temporal Deep Learning Approach for Neonatal Postoperative Pain Assessment.” Computers in Biology and Medicine 129 (August 2020). https://doi.org/10.1016/j.compbiomed.2020.104150.

[16] Mingjing Yang, Huiru Zheng, Haiying Wang, Sally McClean, Jane Hall, and Nigel Harris. 2012. “A Machine Learning Approach to Assessing Gait Patterns for Complex Regional Pain Syndrome.” Medical Engineering and Physics 34 (6): 740–46. https://doi.org/10.1016/j.medengphy.2011.09.018.

[17] Tae Keun Yoo, Sung Kean Kim, Soo Beom Choi, Deog Young Kim, and Deok Won Kim. 2013. “Interpretation of Movement during Stair Ascent for Predicting Severity and Prognosis of Knee Osteoarthritis in Elderly Women Using Support Vector Machine.” Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 192–96. https://doi.org/10.1109/EMBC.2013.6609470.

[18] Naifu Jiang, Keith Dip Kei Luk, and Yong Hu. 2017. “A Machine Learning-Based Surface Electromyography Topography Evaluation for Prognostic Prediction of Functional Restoration Rehabilitation in Chronic Low Back Pain.” Spine 42 (21): 1635–42. https://doi.org/10.1097/BRS.0000000000002159.

[19] Michael JL Sullivan, Pascal Thibault, André Savard, Richard Catchlove, John Kozey, and William D. Stanish. 2006. ”The influence of communication goals and physical demands on different dimensions of pain behavior.” Pain 125, no. 3: 270-277.

[20] Nadia Bianchi-Berthouze, Michel Valstar, Amanda Williams, Joy Egede, Temitayo Olugbade, Chongyang Wang, Hongyin Meng, Min Aung, Nicholas Lane, and Siyang Song. 2020. ”Emopain challenge 2020: Multimodal pain evaluation from facial and bodily expressions.” IEEE International Conference on Automatic Face and Gesture Recognition (FG).

[21] Md Taufeeq Uddin and Shaun Canavan. 2020. “Multimodal Multilevel Fusion for Sequential Protective Behavior Detection and Pain Estimation.” Proceedings - 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020, 844–48. https://doi.org/10.1109/FG47880.2020.00073.

[22] Chongyang Wang, Amanda C De C Williams, Nicholas D Lane, and Nadia Bianchi-berthouze. 2021. “Leveraging Activity Recognition to Enable Protective Behavior Detection in Continuous Data.” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5 (2): 1–24.

[23] Nigel Bosch, Sidney D’Mello, Ryan Baker, Jaclyn Ocumpaugh, Valerie Shute, Matthew Ventura, Lubin Wang, and Weinan Zhao. 2015. “Automatic Detection of Learning-Centered Affective States in the Wild.” International Conference on Intelligent User Interfaces, Proceedings IUI 2015-Janua: 379–88. https://doi.org/10.1145/2678025.2701397.

[24] Cheng Chang, Lei Chen, Cheng Zhang, and Yang Liu. 2018. “An Ensemble Model Using Face and Body Tracking for Engagement Detection.” ICMI 2018 - Proceedings of the 2018 International Conference on Multimodal Interaction, 616–22. https://doi.org/10.1145/3242969.3264986.

[25] Xiaoyang Ma, Min Xu, Yao Dong, and Zhong Sun. 2021. “Automatic Student Engagement in Online Learning Environment Based on Neural Turing Machine.” International Journal of Information and Education Technology 11 (3): 107–11. https://doi.org/10.18178/ijiet.2021.11.3.1497.

[26] Janez Zaletelj, and Andrej Košir. 2017. “Predicting Students’ Attention in the Classroom from Kinect Facial and Body Features.” EURASIP Journal on Image and Video Processing 2017 (1): 80. https://doi.org/10.1186/s13640-017-0228-8.

[27] Chen Sun, Fan Xia, Ye Wang, Yan Liu, Weining Qian, and Aoying Zhou. 2018. “A Deep Learning Model for Automatic Evaluation of Academic Engagement.” Proceedings of the 5th Annual ACM Conference on Learning at Scale, L at S 2018. https://doi.org/10.1145/3231644.3231689.

[28] Erica Volta, Radoslaw Niewiadomski, Temitayo Olugbade, Carla Gilio, Elena Cocchi, Nadia Berthouze, Monica Gori, and Gualtiero Volpe. 2019. “Analysis of Cognitive States during Bodily Exploration of Mathe-

matical Concepts in Visually Impaired Children.” 2019 8th International Conference on Affective Computing and Intelligent Interaction, ACII 2019, 420–26. https://doi.org/10.1109/ACII.2019.8925521.

[29] Simon Senecal, Niels A. Nijdam, and Nadia Magnenat Thalmann. 2019. “Classification of Salsa Dance Level Using Music and Interaction Based Motion Features.” VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications 1: 100–109. https://doi.org/10.5220/0007399701000109.

[30] Saeka Furuichi, Kazuki Abe, and Satoshi Nakamura. 2019. “The Possibility of Personality Extraction Using Skeletal Information in Hip-Hop Dance by Human or Machine.” In INTERACT, edited by David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris, 511–19. Cham: Springer International Publishing.

[31] Camurri, Antonio, Ingrid Lagerlöf, and Gualtiero Volpe. 2003. “Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques.” International Journal of Human Computer Studies 59 (1–2): 213–25. https://doi.org/10.1016/S1071-5819(03)00050-8.

[32] The Pain Consortium. 2016. “UK Pain Messages.” Pain News, 2016.

[33] Singh, Aneesha, Nadia Bianchi-Berthouze, and Amanda CdeC Williams. 2017. “Supporting Everyday Function in Chronic Pain Using Wearable Technology.” In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 3903–15. https://doi.org/10.1145/3025453.3025947.

[34] Temitayo A Olugbade, Aneesha Singh, Nadia Bianchi-Berthouze, Nicolai Marquardt, Min S H Aung, and Amanda Williams. 2019. “How Can Affect Be Detected and Represented in Technological Support for Physical Rehabilitation?” Transactions on Computer-Human Interaction 26 (1): 1–29.

[35] Temitayo Olugbade, Luigi Cuturi, Giulia Cappagli, Paolo Alborno, Joseph Newbold, Nadia Bianchi-berthouze, Gabriel Baud-bovy, Gualtiero Volpe, and Monica Gori. 2017. “What Cognitive and Affective States Should Technology Monitor to Support Learning?” In 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education (MIE’17), 22–25. https://doi.org/https://doi.org/10.1145/3139513.3139522.