Special Issue on New Frontiers in Multimedia-Based and Multimodal HCI

Alessandra Melonio[1] · Maria De Marsico[2] · Cristina Gena[3] · Rosella Gennari[4]

The special issue on "New Frontiers in Multimedia-based and Multimodal HCI" considers research in the area of multimedia-based and multimodal systems for advanced human-computer interaction (HCI) for exploring and expanding on its frontiers. The special issue was inspired by the 2021 Edition of CHItaly, the biannual Conference of the Italian SIGCHI Chapter, held from 11 to 13 July 2021 in Bolzano, Italy. This is an active community that aims at advancing knowledge in the design, development, and application of novel methods, tools, and techniques for HCI. The special issue was promoted through an open call. Therefore, it includes a selection of extended versions of the best papers presented at the conference, besides other high-quality contributions. All submitted manuscripts underwent the same review process, with at least three independent reviewers assessing their relevance, significance, technical quality, and scholarship. The finally accepted papers are 11 out of a total of 28 submissions.

Papers in this special issue consider four main topics, relevant to multimedia and multimodal HCI, and fall into four groups. The first group of papers deals with emotions: how HCI tools can help detect and recognise them, or, on the contrary, how they can cause negative emotional responses. Four papers belong to this group. The three papers in the second group are related to learning, specifically to the design or evaluation of interactive

✉ Alessandra Melonio
alessandra.melonio@unive.it

Maria De Marsico
demarsico@di.uniroma1.it

Cristina Gena
cgena@di.unito.it

Rosella Gennari
gennari@inf.unibz.it

[1] Ca' Foscari University of Venice, Via Torino, 155, 30172, Venice, Italy

[2] Sapienza University of Rome, Via Salaria 113, 00198, Roma, Italy

[3] University of Turin, Via Pessinetto 12, 10149, Torino, Italy

[4] Free University of Bozen-Bolzano, Piazza Domenicani 3, 39100, Bolzano, Italy

technology for learning. The third group gathers two papers that consider how to evaluate and sustain play for social and health purposes. The two papers in the last group, instead, deal with the analysis of human actions in videos and of speech-based interactions in Persian while driving. The contributions of the papers to a topic are briefly outlined in the remainder of this editorial.

1 Emotions and HCI tools

The paper by Giovanna Castellano, Berardina De Carolis, and Nicola Macchiarulo presents a Facial Expression Recognition (FER) system for recognising emotions through facial expressions when faces are masked. In the case of mask detection, the FER system extracts the eye region and recognises the emotion only considering that portion of the face. The effectiveness of the developed FER system was tested: emotions that are related mainly to the mouth region (e.g., disgust) are barely recognised, while positive emotions are well identified by considering only the eye region. Moreover, the authors compared the results of the FER system to the human annotation of emotions on masked faces. Results showed that the FER system outperforms the human annotation.

Another paper, by Rita Francese and Pasquale Attanasio, presents a mobile application able to capture patients' responses to the Beck Depression Inventory-II (BDI-II) questionnaire, together with other data related to patients. The application displays a correlation between the patient's emotional scores and DBI-II scores to the clinician, indicating the relationship between the patient's emotional state and the depression screening score. The authors also conducted a preliminary evaluation involving clinicians and patients to assess the acceptability of the proposed application for its adoption in clinics, besides the user experience (UX) of patients. Results confirm the acceptability of the app in clinical practice, and good results in terms of UX.

People with NDD (neurodevelopmental disorders) tend to have problems with the communication of emotions; the paper by Fabio Catania and Franca Garzotto deals with a conversational agent for emotion expression stimulation in persons with NDD. It presents Emoty, a speech-based conversational agent designed for people with NDD, to train their emotional communication skills. Emoty exploits the emotional expression power of the voice, and hence it engages users in small conversations during which they are asked to repeat sentences and express specific emotions using the appropriate vocal tone. Emoty was evaluated in five individual sessions over two and a half months. The results of the evaluation showed that most users correctly performed the assigned tasks with Emoty, and their capability of expressing emotions with voice consistently improved.

The paper by Antonella De Angeli, Mattia Falduti, María Menéndez Blanco, and Sergio Tessaris deals with the increasing problems raised by non-consensual pornography. The paper analyses the state of the art of interfaces for reporting non-consensual pornography, looking at the effectiveness of different interaction styles from the victim's perspective. The results highlight that the abuse could be very difficult to notify, and the notification can expose the victim to further distress. Moreover, the paper argues that, when the content is spread on several platforms, victims have to move through several interfaces, possibly difficult to find and use, and also likely to subject them to secondary victimisation.

2 Interactive tools for learning

Engagement is a debated construct in education, to the point that there is great variety in the ways engagement is defined and hence assessed. However, there tends to be consensus on the relevance of maintaining engagement in competence-relevant activities, such as learning tasks. The first paper in this group, by Paolo Buono, Berardina Decarolis, Francesca D'Errico, Nicola Macchiarulo, and Giuseppe Palestra, deals with the assessment of student engagement from facial behavior in on-line learning. The paper discusses a system to measure students' engagement in the context of online learning from continuous video streaming. The authors propose and assess a model to automatically predict a student's engagement from facial behavior. Their results show that, globally, engagement prediction from students' facial behavior was weakly correlated to their subjective answers, the analysis of facial movements was positively correlated with the emotional dimension of engagement, while there is an inverse correlation with the gaze, meaning that the more the student feels engaged the less are gaze movements.

The second paper in this group, by Pierpaolo Vittorini and Alessandra Galassi, is about an intelligent tool for data science assignments. The paper describes how the intelligent tool was designed and used in a course about Data Science, within a medical school. The tool supports students in completing data science-related assignments in the R language. Three versions of the tool have been developed and evaluated within two studies. Results indicate the advantages of using the tool in terms of engagement and learning outcomes, suggesting the effectiveness of adopting it as a formative assessment instrument.

The last paper in this group considers another relevant type of learning, namely social-emotional learning for children. The paper, by Alejandro Catala, Hannie Gijlers, and Iris Visser, addresses the problem of designing a storytelling intervention for promoting the social-emotional development of kindergartners. The authors use an emotionally laden story as content and embed a guidance method that can be implemented with either a human or robot guide to enhance the learning setting. The paper presents a study with 70 4–6 years old kindergarteners comparing two guided storytelling activities, guided by a teacher and an autonomous agent, and a non-guided activity. The study assessed children's enactment process, emotion recognition, and story recall. The results show that both guided activities, teacher and agent, did not significantly differ in performance and that the two conditions outperformed the non-guided activity. The paper discusses the results of the study, relevant for further work in the area of social-emotional learning.

3 Playing games for social and health purposes

Both papers in this group deal with games and how to evaluate them. The first paper, by Monica Sanchez de Francisco, Paloma Diaz, Teresa Onorati, and Ignacio Aedo, reports a framework to evaluate multimedia technology in their context of use. The authors tested the proposed framework to evaluate an Augmented Reality app for smartphones and assess if it could be useful to stimulate interest and awareness about the specific urban area where it is used. The presented results are based on a field study exploiting an AR game recreating a historical fact and involving university students.

The second paper, by Paula Alexandra Silva and Renato Santos, investigates how to sustain interest in playing games for health so as to preserve the player's health benefits. The authors introduce a methodology that analyses user comments posted on YouTube about the Just Dance game, to automatically extract information about Usability, User Experience,

and Perceived Health Impacts related to Quality of Life. In doing so, the authors take stock of the available information on public fora and pave the way for using natural language processing on social media data to extract knowledge about the quality-of-life implications of using a game for health. The results of the information extraction and processing can be useful to understand the opinions and impacts perceived by users and to inform the (re)design of games for health.

4 Analysis of human actions in videos and of speech-based interactions

The paper, by Longshuai Sheng and Ce Li, tackles the problem of human action segmentation in video analysis for HCI application. Accurately detecting the category and start time of actions in videos is challenging, due to the lack of large amounts of detailed as well as accurate annotations pertaining to all possibly relevant information in existing datasets. The paper proposes to use the Viterbi algorithm to generate the initial and coarse action segmentation as the baseline and then design a coarse-to-fine learning framework to refine the length partition. The candidate frames of the initial segmentation points are connected in an orderly fashion to construct a fully connected directed graph, and a new coarse-to-fine loss function underlies learning the scores of valid and invalid segmentation paths, aiming at reducing the weight of the scores of invalid segmentation paths and obtain the best video segmentation. Compared with the state-of-the-art methods, the experiments on two popular datasets show that the fine partition model and coarse-to-fine loss function can obtain higher frame accuracy and significantly reduce the time spent on human action segmentation in HCI videos.

The other paper in this group, by Fateme Nazari, Shima Tabibian, and Elaheh Homayounvala, tackles the problem of hands-free communication with in-car equipment. The use of a large number of buttons or a touch screen increases the need for the driver's visual attention and challenges the concentration of drivers on the primary task of driving. A possible partial solution that is quite popular at present is speech-based interaction. The authors present a multimodal user interface design based on touch and speech modes, in Persian, for controlling in-car radio, CD player or music player, fan, heater, and driver-side window.

We close this editorial by thanking all authors for their contributions, and reviewers who considerably helped improve on the submitted manuscripts. We hope readers will enjoy this special issue. The guest Editors Alessandra Melonio, Maria De Marsico, Cristina Gena and Rosella Gennari.